On Databases with Incomplete Information

WITOLD LIPSKI, JR.

Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland

ABSTRACT. Semantic and logical problems arising in an incomplete information database are investigated. A simple query language is described, and its semantics, which refers the queries to the information about reality contained in a database, rather than to reality itself, is defined. This approach, called the internal interpretation, is shown to lead in a natural way to the notions of a topological Boolean algebra and a modal logic related to S4 in the same way as referring queries directly to reality (external interpretation) leads to Boolean algebras and classical logic. An axiom system is given for equivalent (with respect to the internal interpretation) transformation of queries, which is then exploited as a basic tool in a method for computing the internal interpretation for a broad class of queries. An interesting special case of the problem of determining the internal interpretation amounts to deciding whether an assertion about reality (a "yes-no" query) is consistent with the incomplete information about reality contained in a database. A solution to this problem, which relies on the classical combinatorial problem of distinct representatives of subsets, is given.

KEY WORDS AND PHRASES: database, incomplete information, query language semantics, implicit information, modal logic, relational model, null values

CR CATEGORIES: 3.70, 4.33, 5.21

1. *Introduction*

For various reasons, the information contained in a real-world database is usually incomplete. This creates a need for developing methods to handle situations where a database does not contain all the information a user would like to know.

This paper follows a previous paper of the author [13], where a simple mathematical model of a database with incomplete information was introduced. This model, called an *information system* (or just *system*), is based on attributes which can take values in specified attribute domains. Information incompleteness means that instead of having a single value of an attribute, we have a subset of the attribute domain, which represents our knowledge that the actual value is one of the values in this subset, though we do not know which one. This extends the idea of Codd's null value [2], corresponding to the case where this subset is the whole attribute domain. A simple query language to communicate with an information system was also described in [13]. This language includes two kinds of queries, *terms* and *formulas* ("yes-no" queries). The expected response to a term is a list of objects with the property

This work was supported in part by the Polish Academy of Sciences under Contract MR I.3. It was also aided through visiting appointments by the Coordinated Science Laboratory, University of Illinois, and by the Laboratory for Computer Science, Massachusetts Institute of Technology.

Some of the results of this paper were presented at the Third International Conference on Very Large Data Bases, Tokyo, Japan, October 1977.

Author's address: Institute of Computer Science, Polish Academy of Sciences, P.O. Box 22, 00-901 Warsaw PKiN, Poland.

expressed by the term, while the response to a formula is a truth value, T (truth) or F (falsity). It was shown in [13] that when the information is incomplete the same query can be interpreted in many different ways, and to understand those differences, two basic interpretations of a query were introduced, the *external* one and the *internal* one. The external interpretation refers the queries directly to the real world modeled in an incomplete way by the system, so that the external interpretation of a term t is the set of objects which *in reality* have property t. Of course, the information contained in the system is, in general, not sufficient to determine this set exactly. However, in [13] we give methods for computing the best possible bounds on the external interpretation of t logically derivable from the system, that is,

(i) $\|t\|_*$, the set of objects for which we can conclude, from the information available in the system, that they are in the external interpretation of t, and
(ii) $\|t\|^*$, the set of objects for which we cannot rule out the possibility that they belong to the external interpretation of t.

In contrast to the external interpretation, the internal one refers the queries to the system's information about the real world, rather than to the world itself, so that the internal interpretation of a term t is the set of objects for which the information contained in the system satisfies the conditions expressed by t.

Although in the present paper we deal almost exclusively with internal interpretation (so that the word "internal" will sometimes be omitted), the results obtained provide solutions to some problems concerning the external interpretation which were left open in [13]. This paper is organized as follows. In Section 2 we give basic definitions and survey those results of [13] which we shall need here. In Section 3 we precisely define the internal interpretation of queries and investigate its basic properties. In Section 4 we give an axiom system for equivalent (with respect to the internal interpretation) transformation of terms. The technique of equivalent transformation of terms is then used as a basic tool in an algorithm for computing the internal interpretation of an arbitrary term in an arbitrary system. We also prove that our axiom system is complete in the usual logical sense, and we explain the relation of the notion of a topological Boolean algebra to our semantics of terms. In Section 5 we consider a sublanguage which seems to be interesting from the practical point of view. Determining the interpretation of a query in this sublanguage is much easier than in the general case. In particular, we are able to find the interpretation of any formula (which we were not able to do for the general language). Our method of computing the interpretation of an arbitrary formula in the sublanguage, which is described in Section 6, has a combinatorial flavor and is related to the classical problem of distinct representatives of subsets [4]. In Section 7 we discuss some alternative approaches to defining the semantics of queries.

2. *Basic Notions*

In this section we give some basic definitions which we shall need in the rest of the paper. Some of them coincide with those in [13], to which the reader is referred for more detail and motivation.

By an *information system* (or a *system* for short) we mean a triple

$$\mathscr{S} = \langle X, (D_i)_{i \in I}, U \rangle,$$

where

(i) X is a finite set of *objects*,
(ii) I is a finite set of *attributes*,

(iii) D is a nonempty set called the *domain of attribute* i,

(iv) U is a function which associates with every attribute i and every $a \in D_i$ a set $U(i, a) \subseteq X$, such that for every $i \in I$,

$$\bigcup\{U(i, a): a \in D_i\} = X. \tag{1}$$

Intentionally, $U(i, a)$ is the set of objects for which attribute i possibly takes value a. According to this interpretation we can determine for every $x \in X$ and every $i \in I$ the set

$$\beta_i(x) = \{a \in D_i: x \in U(i, a)\} \tag{2}$$

of all possible values attribute i can take for object x. Conversely, U can be obtained from functions β_i, $i \in I$, by the formula

$$U(i, a) = \{x \in X: a \in \beta_i(x)\}. \tag{3}$$

We always assume that the set X of objects, the set I of attributes, and the attribute domains D_i, $i \in I$, are fixed, and we often represent a system by functions β_i, $i \in I$, rather than by U.

Notice that a system represented by β_i, $i \in I$, may be treated as a relational model [1] with only one relation. However, in our case this relation consists of tuples $\langle A_1, \ldots, A_n\rangle$, where each A_i is a *subset* of D_i rather than an element of D_i $(A_i = \beta_i(x))$.

For two systems $\mathscr{S}_1 = \langle X, (D_i)_{i\in I}, U_1\rangle$ and $\mathscr{S}_2 = \langle X, (D_i)_{i\in I}, U_2\rangle$, we say that $\mathscr{S}_2$ is an *extension* of $\mathscr{S}_1$ (in symbols, $\mathscr{S}_1 \preceq \mathscr{S}_2$ or $\mathscr{S}_2 \succeq \mathscr{S}_1$) if

$$U_2(i, a) \subseteq U_1(i, a) \qquad \text{for all} \quad i \in I \quad \text{and} \quad a \in D_i, \tag{4}$$

or equivalently,

$$\beta_i^2(x) \subseteq \beta_i^1(x) \qquad \text{for all} \quad i \in I \quad \text{and} \quad x \in X, \tag{5}$$

where β_i^1 and β_i^2, $i \in I$, correspond to $\mathscr{S}_1$ and $\mathscr{S}_2$, respectively.

Intuitively, $\mathscr{S}_1 \preceq \mathscr{S}_2$ means that the knowledge represented in $\mathscr{S}_1$ is contained in the knowledge represented in $\mathscr{S}_2$. Of course, $\preceq$ is a partial order. A system is called *complete* if $U(i, a) \cap U(i, b) = \emptyset$ for all $a, b \in D_i$, $a \neq b$, or, equivalently, if $\beta_i(x)$ consists of a single value for all $i \in I$, $x \in X$. For a theory of complete systems see [14]. A complete extension is called a *completion.*

Our query language consists of *terms* and *formulas.* Terms are built up from certain elementary parts called *descriptors* and from constants 0, 1 by means of symbols for Boolean operations $-$, $+$, $\cdot$, $\rightarrow$, and a special unary operation $\boxdot$. Every descriptor is of the form $\langle i, A\rangle$, where $i \in I$ and $A \subseteq D_i$; more exactly, A is in a fixed Boolean algebra $\mathscr{B}_i$ of subsets of D_i. Descriptors will be informally written as $\langle$LENGTH $\geq$ 50$\rangle$, $\langle$SEX = M$\rangle$, $\langle$SAL $<$ 10000$\rangle$, $\langle$COLOR = RED$\rangle$ or simply *red*, etc. An example of a term is

$$-(\langle \text{DEPT\#} = 2\rangle \cdot \langle \text{NAME} \neq \text{SMITH}\rangle + \langle \text{SAL} < 50000\rangle \cdot \langle \text{AGE} < 30\rangle).$$

The set of terms is denoted by $\mathscr{T}$.

Formulas are built up from *atomic formulas* $t = s$, where $t, s \in \mathscr{T}$, and from logical constants T, F by means of logical connectives $\neg$, $\vee$, $\wedge$, $\Rightarrow$, and a special (modal) unary connective $\Box$. Finite disjunctions and conjunctions are abbreviated as $\bigvee_{j\in J} \Phi_j$ and $\bigwedge_{j\in J} \Phi_j$, respectively; $\neg(t = s)$ is denoted by $t \neq s$. An example of an (atomic) formula is $(\langle \text{SEX} = \text{F}\rangle \cdot -\langle \text{SAL} > 10000\rangle) = 0$. The set of all formulas is denoted by $\mathscr{F}$. Intuitively, both $\boxdot$ and $\Box$ mean "in every possible extension of our

present knowledge." A query (term or formula) is *simple* if it contains neither ⊡ nor □.

The interpretation of a query Q in a *complete* system $\mathscr{S}$, called the *value* of Q in $\mathscr{S}$ and denoted by $\|Q\|_{\mathscr{S}}$ or just $\|Q\|$, is defined in the natural way. We put

$$\|\langle i, A\rangle\| = \bigcup_{a \in A} U(i, a) = \{x \in X : \beta_i(x) \subseteq A\}, \tag{6}$$

and we interpret 0, 1, −, +, ·, → as ∅, X, and the set-theoretical operations of complementation, union, intersection, and the operation "$(X \setminus A) \cup B$," respectively; for formulas we define $\|t = s\| = T$ iff $\|t\| = \|s\|$, and we interpret the logical connectives in the natural way. Symbols ⊡ and □ are simply disregarded (there can be no proper extension of a complete system). Two queries Q_1 and Q_2 are *externally equivalent* (in symbols, $Q_1 \approx_e Q_2$) if $\|Q_1\|_{\mathscr{S}} = \|Q_2\|_{\mathscr{S}}$ for every *complete* system $\mathscr{S}$ (here D_i, $i \in I$, are fixed, but X and U are arbitrary). In [13] we gave an axiom system **B** for externally equivalent transformation of queries, consisting of the following axiom groups:

B1. Substitutions of terms into the axioms of Boolean algebra, and the axioms of equality (see, e.g., [15]).

B2. The following axioms concerning descriptors:

(i) $\langle i, \varnothing\rangle = 0, \quad \langle i, D_i\rangle = 1,$

(ii) $\langle i, A\rangle + \langle i, B\rangle = \langle i, A \cup B\rangle,$

(iii) $\langle i, A\rangle \cdot \langle i, B\rangle = \langle i, A \cap B\rangle,$

(iv) $-\langle i, A\rangle = \langle i, D_i \setminus A\rangle,$

for all $i \in I$ and all $A, B \in \mathscr{B}_i$.

B3. Substitutions of formulas into the propositional calculus axioms.

B4. The axioms $\boxdot t = t$ for every $t \in \mathscr{T}$ and $\Box\Phi = \Phi$ for every $\Phi \in \mathscr{F}$.

(In fact, B4 is missing in [13], where we consider only simple queries.) It can easily be proved that the axiom system **B** is *complete*, that is, $Q_1 \approx_e Q_2$ iff Q_1 can be transformed into Q_2 by using axioms in **B**. Our main task in this paper is to generalize the definition of the value of a query to arbitrary (not necessarily simple) queries and to arbitrary (not necessarily complete) systems, and then to axiomatize this extended notion of value in the same way as the value of queries in complete systems is axiomatized by the axiom system **B**. We begin this program by giving, in the next section, a precise definition of the value $\|Q\|_{\mathscr{S}}$ of an arbitrary query Q in an arbitrary system $\mathscr{S}$.

3. *Internal Interpretation of Queries*

Before introducing the formal definition of $\|Q\|_{\mathscr{S}}$ for arbitrary Q and $\mathscr{S}$, let us first give some intuitive ideas connected with it. Our definition is based on a consistent approach where any query Q is interpreted as expressing an internal (with respect to the database) property of objects (if Q is a term) or a property of the database as a whole (if Q is a formula); in other words, a term or formula expresses some conditions on the information available about an object or about the whole collection of objects, respectively. Any descriptor $\langle i, A\rangle$ will be understood as "known to have the value of attribute i in A," and the symbols −, +, and · will be interpreted as the usual set-theoretical operations of complementation, union, and intersection, respectively, in exactly the same way as in the case of complete information. Notice that, in particular, the interpretation of *red* + *blue* is "known to be red or known to be blue" (rather

than "known to be [red or blue]"); similarly, $-red$ is interpreted as "not known to be red" (rather than "known not to be red").

The interpretation of $\boxdot t$ will be, roughly speaking, the set of all objects *not only having (internal) property t now, but also in every—not necessarily complete—conceivable extension of our present knowledge.* The interpretation of $\Box\Phi$ is similar—it is T if and only if the assertion concerning the information on our collection of objects, expressed by Φ, is bound to remain true in every possible extension of our present knowledge. It should be emphasized that our interpretation of queries is intended for a user who is fully aware of the fact that the information available in the system may be incomplete, and who may explicitly refer to this incompleteness in his queries (by using $\boxdot$ and $\Box$).

Definition 3.1. Let $\mathscr{S} = \langle X, (D_i)_{i\in I}, U\rangle$ be an arbitrary query. The *value* of Q in $\mathscr{S}$, denoted by $\|Q\|_{\mathscr{S}}$ (or $\|Q\|$ when $\mathscr{S}$ is understood), is defined inductively as follows:

(i) $\|\langle i, A\rangle\| = \{x \in X : \beta_i(x) \subseteq A\}$;
(ii) $\|0\| = \varnothing$, $\|1\| = X$;
(iii) $\|-t\| = X \setminus \|t\|$;
(iv) $\|t + s\| = \|t\| \cup \|s\|$;
(v) $\|t \cdot s\| = \|t\| \cap \|s\|$;
(vi) $\|t \to s\| = (X \setminus \|t\|) \cup \|s\|$;
(vii) $\|\boxdot t\|_{\mathscr{S}} = \{x \in X : \text{for every } \mathscr{S}' \geq \mathscr{S},\ x \in \|t\|_{\mathscr{S}'}\} = \bigcap_{\mathscr{S}'\geq\mathscr{S}} \|t\|_{\mathscr{S}'}$;
(viii) $\|F\| = F$, $\|T\| = T$;
(ix) $\|t = s\| = \begin{cases} T & \text{if } \|t\| = \|s\|, \\ F & \text{otherwise;} \end{cases}$
(x) $\|\neg\Phi\| = \neg\|\Phi\|$;
(xi) $\|\Phi \vee \Psi\| = \|\Phi\| \vee \|\Psi\|$;
(xii) $\|\Phi \wedge \Psi\| = \|\Phi\| \wedge \|\Psi\|$;
(xiii) $\|\Phi \Rightarrow \Psi\| = \neg\|\Phi\| \vee \|\Psi\|$;
(xiv) $\|\Box\Phi\|_{\mathscr{S}} = \begin{cases} T & \text{if } \|\Phi\|_{\mathscr{S}'} = T \text{ for every } \mathscr{S}' \geq \mathscr{S}, \\ F & \text{otherwise} \end{cases}$
$= \inf\{\|\Phi\|_{\mathscr{S}'} : \mathscr{S}' \geq \mathscr{S}\}.$

(inf refers to the natural ordering $F < T$.)

It will be convenient to denote $-\boxdot -t$ by $\diamondsuit t$, for any term t, and $\neg\Box\neg\Phi$ by $\diamondsuit\Phi$, for any formula Φ. $\diamondsuit t$ and $\diamondsuit\Phi$ have a natural interpretation, given by the following theorem.

THEOREM 3.1

(*a*) *For any term t and any system* $\mathscr{S}$,

$$\|\diamondsuit t\|_{\mathscr{S}} = \{x \in X : \textit{for some } \mathscr{S}' \geq \mathscr{S},\ x \in \|t\|_{\mathscr{S}'}\} = \bigcup_{\mathscr{S}'\geq\mathscr{S}} \|t\|_{\mathscr{S}'}.$$

(*b*) *For any formula* Φ *and any system* $\mathscr{S}$,

$$\|\diamondsuit\Phi\|_{\mathscr{S}} = \begin{cases} T & \textit{if for some } \mathscr{S}' \geq \mathscr{S},\ \|\Phi\|_{\mathscr{S}'} = T, \\ F & \textit{otherwise} \end{cases}$$
$$= \sup\{\|\Phi\|_{\mathscr{S}'} : \mathscr{S}' \geq \mathscr{S}\}.$$

PROOF

$$\text{(a)}\quad \|\Diamond t\|_{\mathscr{S}} = \|-\Box-t\|_{\mathscr{S}} = X\setminus \bigcap_{\mathscr{S}'\geq\mathscr{S}} \|-t\|_{\mathscr{S}'} = X\setminus \bigcap_{\mathscr{S}'\geq\mathscr{S}} (X\setminus\|t\|_{\mathscr{S}'})$$
$$= X\setminus(X\setminus \bigcup_{\mathscr{S}'\geq\mathscr{S}} \|t\|_{\mathscr{S}'}) = \bigcup_{\mathscr{S}'\geq\mathscr{S}} \|t\|_{\mathscr{S}'}$$
$$= \{x\in X: \text{for some } \mathscr{S}'\geq\mathscr{S},\ x\in\|t\|_{\mathscr{S}'}\}.$$

$$\text{(b)}\quad \|\Diamond\Phi\|_{\mathscr{S}} = \|\neg\Box\neg\Phi\|_{\mathscr{S}} = \neg\|\Box\neg\Phi\|_{\mathscr{S}}$$
$$= \neg\inf\{\|\neg\Phi\|_{\mathscr{S}'}: \mathscr{S}'\geq\mathscr{S}\} = \neg\neg\sup\{\|\Phi\|_{\mathscr{S}'}: \mathscr{S}'\geq\mathscr{S}\}$$
$$= \sup\{\|\Phi\|_{\mathscr{S}'}: \mathscr{S}'\geq\mathscr{S}\}$$
$$= \begin{cases} T & \text{if for some } \mathscr{S}'\geq\mathscr{S},\ \|\Phi\|_{\mathscr{S}'} = T,\\ F & \text{otherwise.}\end{cases}$$

□

It should be emphasized that in Theorem 3.1 as well as in Definition 3.1 (vii) and (xiv), $\mathscr{S}'$ is *not* assumed to be complete. If an operation $\boxdot'$ were defined to be like $\boxdot$, except that "for every $\mathscr{S}'\geq\mathscr{S}$" is replaced by "for every complete $\mathscr{S}'\geq\mathscr{S}$," then $\boxdot'$ and $\boxdot$ would not coincide. We have, for instance, $\|\boxdot'(\langle \text{SEX} = \text{M}\rangle + \langle \text{SEX} = \text{F}\rangle)\|_{\mathscr{S}} = X$, since in every completion $\mathscr{S}'$ of $\mathscr{S}$, $\|\langle \text{SEX} = \text{M}\rangle + \langle \text{SEX} = \text{F}\rangle\|_{\mathscr{S}'} = \|1\|_{\mathscr{S}'} = X$. On the other hand, if the value of SEX for an object x is not known in $\mathscr{S}$, then

$$x\notin \|\langle \text{SEX} = \text{M}\rangle + \langle \text{SEX} = \text{F}\rangle\|_{\mathscr{S}} = \|\langle \text{SEX} = \text{M}\rangle\|_{\mathscr{S}} \cup \|\langle \text{SEX} = \text{F}\rangle\|_{\mathscr{S}},$$

and consequently $x\notin\|\boxdot(\langle \text{SEX} = \text{M}\rangle + \langle \text{SEX} = \text{F}\rangle)\|_{\mathscr{S}}$ (notice that one of the extensions of $\mathscr{S}$ is $\mathscr{S}$ itself). We see in the next theorem that $\boxdot'$ can be expressed by $\boxdot\Diamond$.

THEOREM 3.2. *For any system $\mathscr{S}$, any term t, and any formula Φ,*

(a) $\|\boxdot\Diamond t\|_{\mathscr{S}} = \{x\in X: \textit{for every completion } \mathscr{S}' \textit{ of } \mathscr{S},\ x\in\|t\|_{\mathscr{S}'}\}$;

(b) $\|\Diamond\boxdot t\|_{\mathscr{S}} = \{x\in X: \textit{for some completion } \mathscr{S}' \textit{ of } \mathscr{S},\ x\in\|t\|_{\mathscr{S}'}\}$;

(c) $\|\Box\Diamond\Phi\|_{\mathscr{S}} = \begin{cases} T & \textit{if for every completion } \mathscr{S}' \textit{ of } \mathscr{S},\ \|\Phi\|_{\mathscr{S}'} = T,\\ F & \textit{otherwise;}\end{cases}$

(d) $\|\Diamond\Box\Phi\|_{\mathscr{S}} = \begin{cases} T & \textit{if there is a completion } \mathscr{S}' \textit{ of } \mathscr{S} \textit{ with } \|\Phi\|_{\mathscr{S}'} = T,\\ F & \textit{otherwise.}\end{cases}$

PROOF. The theorem follows from the structure of the partial order $\preceq$, more specifically, from the fact that for any $\mathscr{S}$ there is a maximal element (i.e., a complete system) $\mathscr{S}'\geq\mathscr{S}$. Let $\alpha(\mathscr{S})$ be an arbitrary assertion with variable $\mathscr{S}$ ranging over systems, which for any particular $\mathscr{S}$ may be either true or false. Then

(i) $(\forall\mathscr{S}'\geq\mathscr{S})(\exists\mathscr{S}''\geq\mathscr{S}')\alpha(\mathscr{S}'')$

is equivalent to

(ii) for every complete $\mathscr{S}'\geq\mathscr{S}$, $\alpha(\mathscr{S}')$.

Indeed, (i) implies

$$(\text{for every complete } \mathscr{S}'\geq\mathscr{S})(\exists\mathscr{S}''\geq\mathscr{S}')\alpha(\mathscr{S}''),$$

which is equivalent to (ii), since for complete $\mathscr{S}'$ the only $\mathscr{S}''\geq\mathscr{S}'$ is $\mathscr{S}'$ itself. Conversely, assume that (ii) holds. Then by the structure of $\preceq$, for every $\mathscr{S}'\geq\mathscr{S}$

there is at least one complete $\mathscr{S}'' \geq \mathscr{S}'$. But $\mathscr{S}'' \geq \mathscr{S}$; hence by (ii), $\alpha(\mathscr{S}'')$ is true, which means that (i) holds. Taking $\alpha(\mathscr{S})$ to be $x \in \|t\|_{\mathscr{S}}$, we obtain

$$\begin{aligned} x \in \|\Box\Diamond t\|_{\mathscr{S}} &\Leftrightarrow (\forall \mathscr{S}' \geq \mathscr{S}) x \in \|\Diamond t\|_{\mathscr{S}'} \\ &\Leftrightarrow (\forall \mathscr{S}' \geq \mathscr{S})(\exists \mathscr{S}'' \geq \mathscr{S}') x \in \|t\|_{\mathscr{S}''} \\ &\Leftrightarrow \text{for every complete } \mathscr{S}' \geq \mathscr{S},\ x \in \|t\|_{\mathscr{S}'}, \end{aligned}$$

which proves (a). Similarly, taking $\alpha(\mathscr{S})$ to be $\|\Phi\|_{\mathscr{S}}$ we have

$$\begin{aligned} \|\Box\Diamond\Phi\|_{\mathscr{S}} &\Leftrightarrow (\forall \mathscr{S}' \geq \mathscr{S})(\exists \mathscr{S}'' \geq \mathscr{S}') \|\Phi\|_{\mathscr{S}''} \\ &\Leftrightarrow \text{for every complete } \mathscr{S}' \geq \mathscr{S},\ \|\Phi\|_{\mathscr{S}'}, \end{aligned}$$

which proves (c).

To prove (b), we have

$$\|\Diamond\Box t\|_{\mathscr{S}} = \|-\Box-\Box t\|_{\mathscr{S}} = \|-\Box-\Box--t\|_{\mathscr{S}} = \|-\Box\Diamond-t\|_{\mathscr{S}} = X\backslash\|\Box\Diamond-t\|_{\mathscr{S}}.$$

Hence, by (a),

$$\begin{aligned} x \in \|\Diamond\Box t\|_{\mathscr{S}} &\Leftrightarrow x \notin \|\Box\Diamond-t\|_{\mathscr{S}} \\ &\Leftrightarrow \neg(\text{for every complete } \mathscr{S}' \geq \mathscr{S},\ x \in \|-t\|_{\mathscr{S}'}) \\ &\Leftrightarrow \text{for some complete } \mathscr{S}' \geq \mathscr{S},\ x \in \|t\|_{\mathscr{S}'}. \end{aligned}$$

Similarly, $\|\Diamond\Box\Phi\| = \neg\|\Box\Diamond\neg\Phi\|$, and we obtain (d) from (c). □

Two queries Q_1, Q_2 are said to be *internally equivalent* (in symbols, $Q_1 \approx_i Q_2$) if $\|Q_1\|_{\mathscr{S}} = \|Q_2\|_{\mathscr{S}}$ for every system $\mathscr{S}$ (as in the definition of external equivalence, D_i, $i \in I$, are fixed, but X and U are arbitrary). It follows trivially from the definition that the internal equivalence is stronger than external, that is, $Q_1 \approx_i Q_2$ implies $Q_1 \approx_e Q_2$. Of course, the converse implication does not hold. For instance, $\langle i, A\rangle + \langle i, B\rangle \approx_e \langle i, A \cup B\rangle$, but in general (if $A \neq A \cup B \neq B$), $\langle i, A\rangle + \langle i, B\rangle \not\approx_i \langle i, A \cup A\rangle$. Indeed,

$$\|\langle i, A\rangle + \langle i, B\rangle\| = \|\langle i, A \cup B\rangle\| = \{x \in X : \beta_i(x) \subseteq A \cup B\},$$

and a subset of $A \cup B$ need not be a subset of either A or B. Another example is the equality $-\langle i, A\rangle = \langle i, D_i\backslash A\rangle$, which is of course true under the external equivalence and is not true (except for trivial cases) under the internal equivalence:

$$\begin{aligned} \|-\langle i, A\rangle\| &= X\backslash\|\langle i, A\rangle\| = X\backslash\{x \in X : \beta_i(x) \subseteq A\} \\ &= \{x \in X : \beta_i(x) \cap (D_i\backslash A) \neq \varnothing\}, \end{aligned}$$

while

$$\|\langle i, D_i\backslash A\rangle\| = \{x \in X : \beta_i(x) \subseteq D_i\backslash A\}.$$

When we put these examples into more concrete terms, $\langle \text{SEX} = \text{M}\rangle + \langle \text{SEX} = \text{F}\rangle$ is interpreted as the set of persons whose sex is known, while $\langle \text{SEX}, \{\text{M}, \text{F}\}\rangle$ is interpreted as the whole set X of persons; similarly, $-\langle \text{SEX} = \text{M}\rangle$ is interpreted as the set of persons who are not known to be men, while $\langle \text{SEX} \neq \text{M}\rangle$ as the set—in general smaller—of persons known not to be men (i.e., known to be women).

4 *Axioms for Internal Interpretation of Terms*

So far we know exactly *what* the internal interpretation of a query *is* (see Definition 3.1), but we do not know *how to compute* it. In this section we develop a method for evaluating the value of an arbitrary term in an arbitrary system. (Formulas are

treated in the next section, where we give simple algorithms for determining the internal interpretation for queries of a special type.)

Our method of computing $\|t\|_{\mathscr{S}}$ is based on transforming t into some equivalent term for which determining $\|\cdot\|_{\mathscr{S}}$ is easy. The transformation process is based on a set of axioms which completely axiomatize the internal equivalence, in the same sense as the axiom system **B** completely axiomatizes the external equivalence.

In order to develop our axiom system, we need some facts about topological Boolean algebras.

Definition 4.1. A *topological Boolean algebra* (TBA for short) is an algebra

$$\langle B, +, \cdot, \rightarrow, -, \mathbb{I}, 0, 1\rangle$$

such that $\langle B, +, \cdot, \rightarrow, -, 0, 1\rangle$ is a Boolean algebra and $\mathbb{I}$ is a unary operation with the following properties:

(i) $\mathbb{I}(a\cdot b) = \mathbb{I}a\cdot\mathbb{I}b$,
(ii) $\mathbb{I}a \leq a$,
(iii) $\mathbb{I}\mathbb{I}a = \mathbb{I}a$,
(iv) $\mathbb{I}1 = 1$,

for all $a, b \in B$ ($a \leq b$ abbreviates $a\cdot b = a$).

An example of a TBA is the Boolean algebra of subsets of a topological space with the operation of taking interior as $\mathbb{I}$. A thorough study of TBAs, as well as the explanation of the elementary topological notions used here, can be found in Rasiowa and Sikorski [15]. The next lemma gives an example of a TBA which plays an important role in our considerations.

LEMMA 4.1. *Let $\langle \mathscr{X}, \leq\rangle$ be a partially ordered set. For all $A, B \subseteq \mathscr{X}$, let*

$$\begin{aligned} \mathbb{I}A &= \{x \in \mathscr{X}: \text{for every } y \geq x,\ y \in A\}, \\ A \Rightarrow B &= (\mathscr{X}\backslash A) \cup B, \\ -A &= \mathscr{X}\backslash A. \end{aligned} \tag{7}$$

Then

$$\langle \mathscr{P}(\mathscr{X}), \cup, \cap, \Rightarrow, -, \mathbb{I}, \varnothing, \mathscr{X}\rangle$$

(where $\mathscr{P}(\mathscr{X})$ denotes the set of all subsets of $\mathscr{X}$) is a TBA.

PROOF. It is sufficient to show that the operation $\mathbb{I}$ defined by (7) satisfies conditions (i)–(iv) of Definition 4.1.

(i) $$\begin{aligned} \mathbb{I}(A \cap B) &= \{x \in \mathscr{X}: (\forall y \geq x) y \in A \cap B\} \\ &= \{x \in \mathscr{X}: (\forall y \geq x) y \geq A\} \cap \{x \in \mathscr{X}: (\forall y \geq x) y \in B\} \\ &= \mathbb{I}A \cap \mathbb{I}B. \end{aligned}$$

(iii) $$\begin{aligned} \mathbb{I}\mathbb{I}A &= \{x \in \mathscr{X}: (\forall y \geq x)(\forall z \geq y) z \in A\} \\ &= \{x \in \mathscr{X}: (\forall z \geq x) z \in A\} = \mathbb{I}A. \end{aligned}$$

(ii) and (iv) are left to the reader. □

Notice that the example provided by this lemma is a special case of the aforementioned general example based on a topological space. Indeed, (7) defines a topological interior operation and hence endows $\mathscr{X}$ with the structure of a topological space. It may be noted that the subsets

$$\{y \in \mathscr{X}: y \geq x\}, \qquad x \in \mathscr{X}$$

form a basis of this topological space.

$\mathbb{I}$ behaves as a topological interior operation, and, similarly, the operation $\mathbb{C}$, defined by

$$\mathbb{C}a = -\mathbb{I}-a,$$

can easily be shown to have the properties of a topological closure operation:

$$\begin{aligned} \mathbb{C}(a+b) &= \mathbb{C}a + \mathbb{C}b, \\ \mathbb{C}\mathbb{C}a &= \mathbb{C}a, \\ a &\leq \mathbb{C}a, \\ \mathbb{C}0 &= 0. \end{aligned}$$

The role of TBAs in the internal interpretation of terms is analogous to that of Boolean algebras in the external interpretation. The analogy is that we can perform internally equivalent transformations of terms using the axioms of TBA, more exactly, the axioms listed below.

Axioms for Terms Under Internal Interpretation. The set **TB** of axioms consists of

TB1. Substitutions of terms into the axioms of TBA, that is, axioms of Boolean algebra, and

(i) $\boxdot(t \cdot s) = \boxdot t \cdot \boxdot s$,
(ii) $t \cdot \boxdot t = \boxdot t$,
(iii) $\boxdot\boxdot t = \boxdot t$,
(iv) $\boxdot 1 = 1$.

TB2. The following axioms concerning descriptors:

(v) $\langle i, \varnothing\rangle = 0$,
(vi) $\langle i, D_i\rangle = 1$,
(vii) $\langle i, A\rangle \cdot \langle i, B\rangle = \langle i, A \cap B\rangle$,

for all $i \in I$, $A, B \in \mathscr{B}_i$.

TB3. The axiom

$$\text{(viii)}\quad \boxdot \sum_{p=1}^{k} \left(-\langle i_p, A_p\rangle + \sum_{q=1}^{n_p} \langle i_p, B_p^q\rangle \right) = \sum_{p=1}^{k} \sum_{q=1}^{n_p} \langle i_p, (D_{ip} \backslash A_p) \cup B_p^q\rangle,$$

for every positive integer k, every sequence of positive integers $n_1, \ldots, n_k$, every sequence of *distinct* attributes $i_1, \ldots, i_k$, and all $A_p, B_p^1, \ldots, B_p^{n_p} \in \mathscr{B}_{i_p}$, $1 \leq p \leq k$.

The last axiom is fairly complicated, but it is hoped that its role will become clear later, when we define the weak multiplicative normal form (see Definition 4.2 below).

Before proving that the axiom set **TB** properly axiomatizes the internal interpretation of terms, we shall give an intuitive explanation why the notion of a TBA is relevant in the context of internal interpretation, more exactly, why transforming terms according to the axioms of TBA preserves internal equivalence. To this end we consider the partially ordered set $\langle \mathscr{X}, \preceq\rangle$, where $\mathscr{X}$ is the set of all systems with fixed X and $(D_i)_{i\in I}$. This partially ordered set defines a TBA

$$\mathscr{A} = \langle \mathscr{P}(\mathscr{X}), \cup, \cap, \Rightarrow, -, \mathbb{I}, \varnothing, \mathscr{X}\rangle \tag{8}$$

(see Lemma 4.1). For any $x \in X$ and $t \in \mathscr{T}$ let

$$f_x(t) = \{\mathscr{S} \in \mathscr{X} : x \in \|t\|_{\mathscr{S}}\}. \tag{9}$$

It will be proved (see Theorem 4.1) that f_x preserves all TBA operations, that is,

$f_x(\boxdot t) = \mathbb{I}f_x(t)$, $f_x(t + s) = f_x(t) \cup f_x(s)$, etc. This means that, loosely speaking, the symbols $\boxdot$, $+, \ldots$ can be interpreted as operations in some TBA. Using this fact, one can easily prove that if $t = s$ can be derived from the axioms of TBA, then $f_x(t) = f_x(s)$ for every $x \in X$, and consequently $\|t\|_{\mathscr{S}} = \|s\|_{\mathscr{S}}$ for every $\mathscr{S}$; that is, $t \approx_i s$.

Let us formulate and prove this more precisely. We write $t \approx_{\mathrm{TB}} s$ if t can be transformed into s by using the axioms in **TB**. Obviously, $\approx_{\mathrm{TB}}$ is an equivalence relation on the set $\mathscr{T}$.

THEOREM 4.1 (ADEQUACY OF THE AXIOM SYSTEM **TB**). *For any terms t, s,*

$$t \approx_{TB} s \qquad \textit{implies} \qquad t \approx_i s.$$

PROOF. It is sufficient to prove that if $t = s$ is an axiom in **TB**, then $t \approx_i s$. To this end, let us consider the partially ordered set $\langle \mathscr{X}, \preceq \rangle$ of all systems with fixed X and $(D_i)_{i \in I}$, ordered by the relation of extension. The set $\mathscr{P}(\mathscr{X})$ of all subsets of $\mathscr{X}$, together with the usual set-theoretical operations and the operation $\mathbb{I}$ defined for every $A \subseteq \mathscr{X}$ by

$$\mathbb{I}A = \{\mathscr{S} \in \mathscr{X} : \text{for every } \mathscr{S}' \geq \mathscr{S}, \mathscr{S}' \in A\},$$

defines a TBA $\mathscr{A}$ (see (8)). For any $x \in X$, let the mapping

$$f_x : \mathscr{T} \to \mathscr{P}(\mathscr{X})$$

be defined by (9). The mapping f_x has the following properties:

$$\begin{aligned} f_x(t + s) &= f_x(t) \cup f_x(s), \\ f_x(t \cdot s) &= f_x(t) \cap f_x(s), \\ f_x(t \to s) &= f_x(t) \Rightarrow f_x(s), \\ f_x(-t) &= -f_x(t), \\ f_x(0) &= \varnothing, \\ f_x(1) &= \mathscr{X}, \\ f_x(\boxdot t) &= \mathbb{I}f_x(t). \end{aligned}$$

We prove the first and the last property. The others are left to the reader.

$$\begin{aligned} f_x(t + s) &= \{\mathscr{S} \in \mathscr{X} : x \in \|t + s\|_{\mathscr{S}}\} \\ &= \{\mathscr{S} \in \mathscr{X} : x \in \|t\|_{\mathscr{S}} \vee x \in \|s\|_{\mathscr{S}}\} \\ &= \{\mathscr{S} \in \mathscr{X} : x \in \|t\|_{\mathscr{S}}\} \cup \{\mathscr{S} \in \mathscr{X} : x \in \|s\|_{\mathscr{S}}\} \\ &= f_x(t) \cup f_x(s). \\ f_x(\boxdot t) &= \{\mathscr{S} \in \mathscr{X} : x \in \|\boxdot t\|_{\mathscr{S}}\} \\ &= \{\mathscr{S} \in \mathscr{X} : (\forall \mathscr{S}' \geq \mathscr{S})\, x \in \|t\|_{\mathscr{S}'}\} \\ &= \{\mathscr{S} \in \mathscr{X} : (\forall \mathscr{S}' \geq \mathscr{S})\, \mathscr{S}' \in f_x(t)\} \\ &= \mathbb{I}f_x(t). \end{aligned}$$

Let $t = s$ be an axiom of group TB1, that is, a substitution of terms into an axiom of TBA. Since this axiom holds true in A, we infer from the properties of f_x that $f_x(t) = f_x(s)$ for every $x \in X$. We illustrate this for axioms $t \cdot (s + r) = t \cdot s + t \cdot r$ and $\boxdot(t \cdot s) = \boxdot t \cdot \boxdot s$:

$$\begin{aligned} f_x(t \cdot (s + r)) &= f_x(t) \cap (f_x(s) \cup f_x(r)) \\ &= (f_x(t) \cap f_x(s)) \cup (f_x(t) \cap f_x(r)) = f_x(t \cdot s + t \cdot r), \\ f_x(\boxdot(t \cdot s)) &= \mathbb{I}(f_x(t) \cap f_x(s)) = \mathbb{I}f_x(t) \cap \mathbb{I}f_x(s) = f_x(\boxdot t \cdot \boxdot s). \end{aligned}$$

Now we prove the same for axioms of group TB2. As (v) and (vi) are trivial, we restrict ourselves to (vii):

$$\begin{aligned} f_x(\langle i, A\rangle \cdot \langle i, B\rangle) &= \{\mathscr{S} \in \mathscr{X} : x \in \| \langle i, A\rangle \cdot \langle i, B\rangle \|_{\mathscr{S}}\} \\ &= \{\mathscr{S} \in \mathscr{X} : x \in \| \langle i, A\rangle \|_{\mathscr{S}} \cap \| \langle i, B\rangle \|_{\mathscr{S}}\} \\ &= \{\mathscr{S} \in \mathscr{X} : \beta_i(x) \subseteq A \wedge \beta_i(x) \subseteq B\} \\ &= \{\mathscr{S} \in \mathscr{X} : \beta_i(x) \subseteq A \cap B\} \\ &= \{\mathscr{S} \in \mathscr{X} : x \in \| \langle i, A \cap B\rangle \|_{\mathscr{S}}\} = f_x(\langle i, A \cap B\rangle). \end{aligned}$$

(Here β_i, $i \in I$, denote the functions uniquely corresponding to $\mathscr{S}$.)

The last axiom to consider is (viii). This is the most difficult part of the theorem.

$$\begin{aligned} & f_x\Big(\boxdot \sum_{p=1}^{k} [-\langle i_p, A_p\rangle + \sum_{q=1}^{n_p} \langle i_p, B_p^q\rangle]\Big) \\ &= \left\{\mathscr{S} \in \mathscr{X} : (\forall \mathscr{S}' \geq \mathscr{S}) x \in \bigcup_{p=1}^{k} \left[\| -\langle i_p, A_p\rangle \|_{\mathscr{S}'} \cup \bigcup_{q=1}^{n_p} \| \langle i_p, B_p^q\rangle \|_{\mathscr{S}'}\right]\right\} \\ &= \{\mathscr{S} \in \mathscr{X} : (\forall \mathscr{S}' \geq \mathscr{S})(\exists p)[x \notin \| \langle i_p, A_p\rangle \|_{\mathscr{S}'} \vee (\exists q) x \in \| \langle i_p, B_p^q\rangle \|_{\mathscr{S}'}]\}. \end{aligned}$$

Taking into account that

$$\begin{aligned} x \notin \| \langle i_p, A_p\rangle \|_{\mathscr{S}} &\Leftrightarrow \neg(A_p \subseteq \beta_{i_p}(x)), \\ x \in \| \langle i_p, B_p^q\rangle \|_{\mathscr{S}} &\Leftrightarrow \beta_{i_p}(x) \subseteq B_p^q, \end{aligned}$$

(β_i, $i \in I$, correspond to $\mathscr{S}$), and using (5), we obtain

$$\begin{aligned} & \mathscr{S} \in f_x\left(\boxdot \sum_{p=1}^{k} \left[-\langle i_p, A_p\rangle + \sum_{q=1}^{n_p} \langle i_p, B_p^q\rangle\right]\right) \\ &\Leftrightarrow (\text{for all nonempty } Z_1 \subseteq \beta_{i_1}(x), \ldots, Z_k \subseteq \beta_{i_k}(x)) \\ &\quad (\exists p)[\neg(Z_p \subseteq A_p) \vee (\exists q) Z_p \subseteq B_p^q] \\ &\Leftrightarrow (\exists p)(\text{for every nonempty } Z \subseteq \beta_{i_p}(x))[Z \subseteq A_p \Rightarrow (\exists q) Z \subseteq B_p^q] \\ &\Leftrightarrow (\exists p)(\text{for every nonempty } Z \subseteq \beta_{i_p}(x) \cap A_p)(\exists q) Z \subseteq B_p^q \\ &\Leftrightarrow (\exists p)(\exists q) \beta_{i_p}(x) \cap A_p \subseteq B_p^q \\ &\Leftrightarrow (\exists p)(\exists q) \beta_{i_p}(x) \subseteq (D_{i_p} \backslash A_p) \cup B_p^q \\ &\Leftrightarrow x \in \bigcup_{p=1}^{k} \bigcup_{q=1}^{n_p} \| \langle i_p, (D_{i_p} \backslash A_p) \cup B_p^q\rangle \|_{\mathscr{S}} \\ &\Leftrightarrow \mathscr{S} \in f_x\left(\sum_{p=1}^{k} \sum_{q=1}^{n_p} \langle i_p, (D_{i_p} \backslash A) \cup B_p^q\rangle\right). \end{aligned}$$

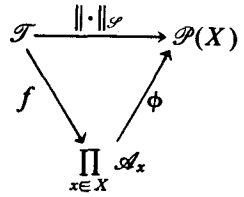
To sum up, we have proved that for every axiom $t = s$ in **TB** and for every $x \in X$, $f_x(t) = f_x(s)$. But this means that for every such axiom, $t \approx_i s$. Indeed,

$$\| t \|_{\mathscr{S}} = \{x \in X : \mathscr{S} \in f_x(t)\} = \{x \in X : \mathscr{S} \in f_x(s)\} = \| s \|_{\mathscr{S}}.$$

Now the theorem follows from the fact that each time we use an axiom $t = s$ in the transformation process, we replace an occurrence of t by s in some term r. Since $t \approx_i s$, this operation does not change the value of r. □

Putting our proof into more algebraic terms, we may illustrate it by the following

commutative diagram:



where $\prod_{x\in X}\mathscr{A}_x$ denotes the direct product of TBAs $\mathscr{A}_x$, each $\mathscr{A}_x$ being a copy of $\mathscr{A} = \langle \mathscr{P}(\mathscr{X}), \cup, \cap, \Rightarrow, -, \mathbb{I}, \varnothing, \mathscr{X}\rangle$, and where $f = (f_x)_{x\in X}$ and

$$\phi(a) = \{x \in X : \mathscr{S} \in a_x\}$$

for every $a = (a_x)_{x\in X} \in \prod_{x\in X}\mathscr{A}_x$. The theorem follows from the fact that f is a homomorphism (i.e., it preserves all TBA operations) and that $\prod_{x\in X}\mathscr{A}_x$ is a TBA.

We prove in this section that the converse of Theorem 4.1 holds as well, which means that the relations $\approx_i$ and $\approx_{\mathrm{TB}}$ coincide (see Theorem 4.4 below).

Using Theorem 4.1, we can obtain some useful corollaries from the axioms in **TB.**

LEMMA 4.2

(a) $\boxdot\langle i, A\rangle \approx_i \langle i, A\rangle$;
(b) $\boxdot -\langle i, A\rangle \approx_i \langle i, D_i\backslash A\rangle$;
(c) $\boxdot(\langle i_1, A_1\rangle + \cdots + \langle i_k, A_k\rangle) \approx_i \langle i_1, A_1\rangle + \cdots + \langle i_k, A_k\rangle$
for arbitrary (not necessarily distinct) $i_1, \ldots, i_k$;
(d) $\boxdot(-\langle i, A_1\rangle + \cdots + -\langle i, A_k\rangle) \approx_i \langle i, D_i\backslash(A_1 \cap \cdots \cap A_k)\rangle$;
(e) $\boxdot(-\langle i_1, A_1\rangle + \cdots + -\langle i_k, A_k\rangle) \approx_i \langle i_1, D_{i_1}\backslash A_1\rangle + \cdots + \langle i_k, D_{i_k}\backslash A_k\rangle$
whenever $i_1, \ldots, i_k$ *are pairwise distinct*;
(f) $\diamondsuit\langle i, A\rangle \approx_i -\langle i, D_i\backslash A\rangle$;
(g) $\diamondsuit -\langle i, A\rangle \approx_i -\langle i, A\rangle$;
(h) $\diamondsuit(\langle i_1, A_1\rangle + \cdots + \langle i_k, A_k\rangle) \approx_i -\langle i_1, D_{i_1}\backslash A_1\rangle \cdot \cdots \cdot -\langle i_k, D_{i_k}\backslash A_k\rangle$
whenever $i_1, \ldots, i_k$ *are pairwise distinct*;
(i) $\diamondsuit(\langle i, A_1\rangle + \cdots + \langle i, A_k\rangle) \approx_i -\langle i, D_i\backslash(A_1 \cup \cdots \cup A_k)\rangle$;
(j) $\diamondsuit(-\langle i_1, A_1\rangle \cdot \cdots \cdot -\langle i_k, A_k\rangle) \approx_i -\langle i_1, A_1\rangle \cdot \cdots \cdot -\langle i_k, A_k\rangle$
for arbitrary (not necessarily distinct) $i_1, \ldots, i_k$.

PROOF. (a) through (e) correspond to some special cases of axiom (viii). For instance,

$$\boxdot -\langle i, A\rangle \approx_i \boxdot(-\langle i, A\rangle + \langle i, \varnothing\rangle) \approx_i \langle i, D_i\backslash A\rangle.$$

(f) through (j) can be derived from (a) through (e). For instance,

$$\begin{aligned}\diamondsuit(\langle i_1, A_1\rangle \cdot \cdots \cdot \langle i_k, A_k\rangle) &\approx_i -\boxdot -(\langle i_1, A_1\rangle \cdot \cdots \cdot \langle i_k, A_k\rangle)\\ &\approx_i -\boxdot(-\langle i_1, A_1\rangle + \cdots + -\langle i_k, A_k\rangle) \approx_i -(\langle i_1, D_{i_1}\backslash A_1\rangle + \cdots + \langle i_k, D_{i_k}\backslash A_k\rangle)\\ &\approx_i -\langle i_1, D_{i_1}\backslash A_1\rangle \cdot \cdots \cdot -\langle i_k, D_{i_k}\backslash A_k\rangle.\end{aligned}$$

□

Now we are prepared to describe a method of determining the value of any term in any system. The main tool in our approach is a certain normal form for terms.

Definition 4.2 (i) A term is *weakly coprimitive* if it is of the form

$$\sum_{p=1}^{k}\left(-\langle i_p, A_p\rangle + \sum_{q=1}^{m}\langle i_p, B_p^q\rangle\right), \tag{10}$$

where the attributes $i_1, \ldots, i_k$ are pairwise distinct.

(ii) A term is in *weak multiplicative normal form* (WMNF) if it is of the form

$$\prod_{k\in K} t_k,$$

where all t_k's are weakly coprimitive.

Notice that any term in WMNF is simple. If we had axioms

$$-\langle i, A\rangle = \langle i, D_i\backslash A\rangle,$$
$$\langle i, A\rangle + \langle i, B\rangle = \langle i, A \cup B\rangle,$$

then we would be able to transform (10) into

$$\sum_{p=1}^{k} \langle i_p, (D_{i_p}\backslash A_p) \cup \bigcup_{q=1}^{m} B_p^q\rangle.$$

However, as we have already seen, these axioms are not valid under the internal interpretation.

LEMMA 4.3. *For any simple term t there is a term s in WMNF such that* $t \approx_{TB} s$.

PROOF. We shall describe an effective algorithm for transforming t into an internally equivalent term s in WMNF. Using the axioms of Boolean algebra, we transform t into a product of sums, each sum consisting of some number of descriptors or negations of descriptors. Each of these sums can be transformed into the form where descriptors are grouped according to attributes, say

$$\sum_{p=1}^{k} \left(\sum_{r=1}^{m_p} -\langle i_p, A_p^r\rangle + \sum_{q=1}^{n_p} \langle i_p, B_p^q\rangle \right) \tag{11}$$

($i_1, \ldots, i_k$ pairwise distinct). By de Morgan's Law and axiom (vii) of group TB2,

$$\sum_{r=1}^{m_p} -\langle i_p, A_p^r\rangle \approx_{\mathrm{TB}} -\prod_{r=1}^{m_p} \langle i_p, A_p^r\rangle \approx_{\mathrm{TB}} -\langle i_p, \bigcap_{r=1}^{m_p} A_p^r\rangle.$$

This enables us to transform (11) into a weakly coprimitive term, and consequently to transform the whole term into WMNF. □

THEOREM 4.2. *For any term t there is a term s in WMNF such that* $t \approx_{TB} s$.

PROOF. First we describe how to eliminate $\boxdot$ from t. If t contains a $\boxdot$, then t must contain a subterm of the form $\boxdot p$ with no $\boxdot$ occurring in p. By the previous lemma, p can be transformed into a term in WMNF, say p'. But axiom (viii) enables us to directly transform $\boxdot p'$ into a term in WMNF. In this way the number of occurrences of $\boxdot$ is decreased by one. By repeating the above procedure we ultimately eliminate $\boxdot$ from t. Now it suffices to transform the resulting simple term into WMNF, which is possible by the previous lemma. □

Notice that the proof of Theorem 4.2 provides an effective algorithm for computing the value of any term in an arbitrary system. Indeed, in order to evaluate $\|t\|$ for a term t not containing $\boxdot$, we can directly apply Definition 3.1(i)–(vi). For general terms, we give the following convenient formulation of our method of determining the internal interpretation.

THEOREM 4.3. *For any term t and any system* $\mathscr{S}$, $x \in \|t\|_{\mathscr{S}}$ *if and only if after transforming t into WMNF, for every factor*

$$\sum_{p=1}^{k} \left(-\langle i_p, A_p\rangle + \sum_{q=1}^{n_p} \langle i_p, B_p^q\rangle \right) \tag{12}$$

of this WMNF there is an attribute i_p such that

$$\beta_{i_p}(x) \not\subseteq A_p \quad \text{or} \quad \text{for some } q,\ \beta_{i_p}(x) \subseteq B_p^q.$$

PROOF. The proof follows directly from Definition 3.1 and from the fact that our process of transforming into WMNF preserves internal equivalence. □

Example 4.1. Let t be the following term:

$$-\langle \text{AGE} < 40\rangle \cdot \boxdot(\langle \text{SAL} > 10000\rangle \cdot -\langle \text{SAL} > 20000\rangle + \langle \text{SEX} = \text{F}\rangle + \langle \text{SEX} = \text{M}\rangle). \tag{13}$$

This query asks for objects with the value of AGE not known to be less than 40 (it may be known to be ≥ 40), which not only now but also in every possible extension of our present knowledge have the following (internal) property: Either the value of SAL is known to be greater than 10000 and is not known to be greater than 20000 (it may be known to be ≤ 20000), or the value of SEX is known. We transform t into WMNF:

$$\begin{aligned} t &\approx_i \langle \text{AGE} < 40\rangle \cdot \boxdot((\langle \text{SAL} < 10000\rangle + \langle \text{SEX} = \text{F}\rangle + \langle \text{SEX} = \text{M}\rangle) \\ &\qquad \cdot(-\langle \text{SAL} > 20000\rangle + \langle \text{SEX} = \text{F}\rangle + \langle \text{SEX} = \text{M}\rangle)) \\ &\approx_i -\langle \text{AGE} < 40\rangle \cdot \boxdot(\langle \text{SAL} > 10000\rangle + \langle \text{SEX} = \text{F}\rangle + \langle \text{SEX} = \text{M}\rangle) \\ &\quad \cdot \boxdot(-\langle \text{SAL} > 20000\rangle + \langle \text{SEX} = \text{F}\rangle + \langle \text{SEX} = \text{M}\rangle) \\ &\approx_i -\langle \text{AGE} < 40\rangle \cdot (\langle \text{SAL} > 10000\rangle + \langle \text{SEX} = \text{F}\rangle + \langle \text{SEX} = \text{M}\rangle) \\ &\quad \cdot(\langle \text{SAL} \leq 20000\rangle + \langle \text{SEX} = \text{F}\rangle + \langle \text{SEX} = \text{M}\rangle). \end{aligned}$$

(It is easy to see that in this particular case t can be further transformed into $-\langle \text{AGE} < 40\rangle \cdot (\langle \text{SAL IN } (10000, 20000]\rangle + \langle \text{SEX} = \text{F}\rangle + \langle \text{SEX} = \text{M}\rangle)$.) Now consider the following system:

object	AGE	SAL	SEX
x_1	$(0, \infty)$	$[0, \infty)$	{M}
x_2	{30}	{20000}	{F}
x_3	$(20, \infty)$	(15000, 30000)	{F, M}
x_4	(35, 45)	{15000}	{F, M}

Using Theorem 4.3, we obtain the value of (13):

$$\|t\| = \{x_1, x_4\}.$$

Note that in some cases it is useful to exploit the fact that

$$\begin{aligned} -\langle i, A\rangle + \langle i, B\rangle &\approx_{\text{TB}} (-\langle i, A\rangle + \langle i, B\rangle)\cdot(-\langle i, A\rangle + \langle i, A\rangle) \\ &\approx_{\text{TB}} -\langle i, A\rangle + -\langle i, A\rangle \cdot \langle i, B\rangle + \langle i, A\rangle \cdot \langle i, B\rangle \\ &\approx_{\text{TB}} -\langle i, A\rangle + \langle i, A \cap B\rangle \end{aligned} \tag{14}$$

for transforming every factor (12) of a WMNF into

$$\sum_{p=1}^{k} \left(-\langle i_p, A_p\rangle + \sum_{q=1}^{n_p} \langle i_p, A_p \cap B_p^q\rangle \right).$$

In this way we can eliminate from (12) the summands $\langle i_p, B_p^q\rangle$ with $A_p \cap B_p^q = \varnothing$.

We can use the technique of transforming into WMNF to prove the following theorem.

THEOREM 4.4 (COMPLETENESS OF THE AXIOM SYSTEM **TB**). *For any terms t, s,*

$$t \approx_i s \quad \text{implies} \quad t \approx_{\text{TB}} s.$$

PROOF. Assume that $t \approx_i s$, and transform $(-t + s)\cdot(-s + t)$ into a term r in WMNF, $r \approx_{\mathrm{TB}} (-t + s)\cdot(-s + t)$. We shall prove that $r \approx_{\mathrm{TB}} 1$. Indeed, otherwise r would contain a factor r_i, say

$$\sum_{p=1}^{k}\left(-\langle i_p, A_p\rangle + \sum_{q=1}^{n_p}\langle i_p, B_p^q\rangle\right),$$

with $A_p \not\subseteq B_p^q$ for $1 \le p \le k$, $1 \le q \le n_p$. Notice that if $A_p \subseteq B_p^q$, then, by (14),

$$\begin{aligned}-\langle i_p, A_p\rangle + \langle i_p, B_p^q\rangle &\approx_{\mathrm{TB}} -\langle i_p, A_p\rangle + \langle i_p, A_p \cap B_p^q\rangle\\ &\approx_{\mathrm{TB}} -\langle i_p, A_p\rangle + \langle i_p, A_p\rangle \approx_{\mathrm{TB}} 1.\end{aligned}$$

Consider a system $\mathscr{S}$ with $\beta_i(x) = D_i$ for all $i \in I$, $x \in X$, and with a nonempty set X of objects. By axiom (viii) we get

$$\begin{aligned}\|\boxdot r_i\|_{\mathscr{S}} &= \left\|\sum_{p=1}^{k}\sum_{q=1}^{n_p}\langle i_p, (D_{i_p}\backslash A_p)\cup B_p^q\rangle\right\|_{\mathscr{S}}\\ &= \bigcup_{p=1}^{k}\bigcup_{q=1}^{n_p}\|\langle i_p, (D_{i_p}\backslash A_p)\cup B_p^q\rangle\|_{\mathscr{S}}\\ &= \bigcup_{p=1}^{k}\bigcup_{q=1}^{n_p}\varnothing = \varnothing,\end{aligned}$$

since $A_p \not\subseteq B_p^q$ implies $(D_{i_p}\backslash A_p)\cup B_p^q \neq D_{i_p}$. Let us fix an object $x \in X$. By the definition of $\|r_i\|_{\mathscr{S}}$ (see Definition 3.1(vii)), there is a system $\mathscr{S}' \ge \mathscr{S}$ such that $x \notin \|r_i\|_{\mathscr{S}'}$, and consequently $x \notin \|r\|_{\mathscr{S}'}$ ($\|r\|_{\mathscr{S}'} \subseteq \|r_i\|_{\mathscr{S}'}$). On the other hand, since $t \approx_i s$, we have $\|t\|_{\mathscr{S}'} = \|s\|_{\mathscr{S}'}$ and

$$\begin{aligned}\|r\|_{\mathscr{S}'} &= \|(-t + s)\cdot(-s + t)\|_{\mathscr{S}'}\\ &= ((X\backslash\|t\|_{\mathscr{S}'})\cup\|s\|_{\mathscr{S}'})\cap((X\backslash\|s\|_{\mathscr{S}'})\cup\|t\|_{\mathscr{S}'})\\ &= X\cap X = X,\end{aligned}$$

that is, $x \in \|r\|_{\mathscr{S}'}$. This contradiction shows that $r \approx_{\mathrm{TB}} 1$ must hold. Hence we have

$$\begin{aligned}t &\approx_{\mathrm{TB}} t\cdot r \approx_{\mathrm{TB}} t\cdot(-t + s)\cdot(-s + t) \approx_{\mathrm{TB}} t\cdot(-t + s)\\ &\approx_{\mathrm{TB}} t\cdot(t + s)\cdot(-t + s) \approx_{\mathrm{TB}} t\cdot(s + (t\cdot -t)) \approx_{\mathrm{TB}} t\cdot s.\end{aligned}$$

Similarly $s \approx_{\mathrm{TB}} t\cdot s$, and consequently $t \approx_{\mathrm{TB}} s$. □

Combining this result with Theorem 4.1 we see that the relations $\approx_i$ and $\approx_{\mathrm{TB}}$ coincide.

In some cases it may be convenient to use another axiom system based on $\diamond$ as a primitive operation ($\boxdot$ can be expressed as $-\diamond-$). The reader may easily verify that a (dual) complete axiom system **TB*** can be obtained from **TB** by replacing axioms (i)–(iv) and (viii) by

(i)* $\diamond(t + s) = \diamond t + \diamond s$,

(ii)* $t\cdot\diamond t = t$,

(iii)* $\diamond\diamond t = \diamond t$,

(iv)* $\diamond 0 = 0$,

(viii)* $$\diamond\prod_{p=1}^{k}\left(\langle i_p, A_p\rangle\cdot\prod_{q=1}^{n_p} -\langle i_p, B_p^q\rangle\right) = \prod_{p=1}^{k}\prod_{q=1}^{n_p} -\langle i_p, (D_{i_p}\backslash A_p)\cup B_p^q\rangle$$

(with the same restrictions as in (viii)).

Using these axioms, we can transform any term into *weak additive normal form* (WANF), that is, a sum of *weakly primitive* terms of the form

$$\prod_{p=1}^{k}\left(\langle i_p, A_p\rangle\cdot\prod_{q=1}^{n_p} -\langle i_p, B_p^q\rangle\right)$$

($i_1, \ldots, i_p$ pairwise distinct). Computing the internal interpretation is then carried out analogously as for WMNF.

Now we briefly discuss the internal interpretation of formulas. We do not give any general method of computing $\|\Phi\|_{\mathscr{S}}$; no such method is known to the author.

A reader who is familiar with the Kripke models for the modal logic S4 (see Kripke [8] and Fitting [3, Ch. 3]) has undoubtedly noticed the similarity between Definition 3.1 and the definition of truth value of a formula in a Kripke model. An immediate corollary from this similarity is that transforming formulas according to the axioms of modal logic S4 preserves internal equivalence. For a discussion of S4 the reader is referred to [3, 6, 8, 15]. Here we only note that TBAs play in S4 the same role as Boolean algebras in the classical logic; that is, an expression is an S4-tautology if and only if its value is 1 in every TBA. An important difference between Definition 3.1 and a Kripke model is that the latter can be based on an arbitrary partial order, while our partial order $\preceq$ (the relation of extension) has some specific properties; for example, for any $\mathscr{S}$ there is a maximal element (complete system) $\mathscr{S}' \succeq \mathscr{S}$. This has the result that there are formulas which have value T in every system, yet which are not (substitutions of formulas into) S4 tautologies. Examples of such formulas are

$$\Box\Diamond\Phi \Rightarrow \Diamond\Box\Phi,$$
$$\Box\Diamond(\Phi \wedge \Psi) \Leftrightarrow \Box\Diamond\Phi \wedge \Box\Diamond\Psi,$$
$$\Diamond\Box(\Phi \vee \Psi) \Leftrightarrow \Diamond\Box\Phi \vee \Diamond\Box\Psi,$$

($\Phi \Leftrightarrow \Psi$ abbreviates $(\Phi \Rightarrow \Psi) \wedge (\Psi \Rightarrow \Phi)$; see Theorem 3.2 for the intuition connected with these formulas). Other universally valid formulas are, in our case,

$$\Box(t = 0) \Leftrightarrow \Diamond t = 0,$$
$$\Box(t \neq 0) \Leftrightarrow \boxdot t \neq 0,$$
$$\Diamond(t = 0) \Leftrightarrow \boxdot t = 0,$$
$$\Diamond(t \neq 0) \Leftrightarrow \Diamond t \neq 0.$$

It is not known to the author whether all these formulas completely axiomatize the internal equivalence of formulas (i.e., whether an analog of Theorem 4.4 holds). While the problem of a complete axiomatization of internal equivalence, as well as that of evaluating $\|\Phi\|_{\mathscr{S}}$ for any Φ and $\mathscr{S}$, remain interesting logical open questions, it seems that the method of determining the internal interpretation for formulas of a special kind which we describe in the next section is quite sufficient for practical purposes.

Let us finally mention that our terms and formulas can be treated as the open and closed formulas, respectively, of a certain monadic modal predicate calculus (see Lipski [12]).

5. *Internal Interpretation: A Simplified Language*

The internal interpretation of queries described in the preceding section, although precisely defined, may be not intuitively clear for a user, especially a casual one. The main reason seems to be the fact that the meaning of the operation $\boxdot$ and the connective $\Box$ is less lucid than the meaning of the (formally more complicated) operation $\boxdot\Diamond$ and connective $\Box\Diamond$ (see Theorem 3.2). A user, who is in most cases interested just in deducing as much information about reality as possible from incomplete data, is likely to think of the system in terms of all completions of the information available in the system, that is, all possibilities of what reality may turn out to look like. On the other hand, in the definition of $\|\boxdot t\|$ and $\|\Box\Phi\|$ (see Definition 3.1(vii) and (xiv)) we take into consideration *all* extensions, not necessarily

complete. Each such extension may be thought of as an intermediate stage in a hypothetical process of increasing the information contained in the system. What we do in the interpretation described in the preceding section is, in a sense, indirectly define (internal) properties of objects (or of a system as a whole) by specifying properties of the possible processes of increasing knowledge. It seems that such an expressive power may not be necessary in a query language. With this in mind, we now propose a certain subclass of queries as a basis for the query language.

Let us denote $\boxdot\diamond t$ by *surely* t, for any term t. The set $\mathscr{T}_0$ of *special terms* is defined to be the least set $\mathscr{T}'$ with the following two properties:

(i) 0, 1, and every descriptor are in $\mathscr{T}'$.
(ii) $-t$, *surely* t, $(t + s)$, $(t \cdot s)$, and $(t \to s)$ are in $\mathscr{T}'$ whenever $t, s \in \mathscr{T}'$.

If we denote $-surely - t$ by *possibly* t, then

$$possibly\ t \approx_i -\boxdot\diamond -t \approx_i -\boxdot -\boxdot --t \approx_i \diamond\boxdot t,$$

and by Theorem 3.2, for any term t we have

$$\|surely\ t\|_{\mathscr{S}} = \bigcap\{\|t\|_{\mathscr{S}'} : \mathscr{S}' \text{ is a completion of } \mathscr{S}\}, \tag{15}$$

$$\|possibly\ t\|_{\mathscr{S}} = \bigcup\{\|t\|_{\mathscr{S}'} : \mathscr{S}' \text{ is a completion of } \mathscr{S}\}. \tag{16}$$

Similarly, by introducing the abbreviations *Surely* Φ for $\Box\diamond\Phi$ and *Possibly* Φ for $\neg Surely \neg\Phi$, we define the set $\mathscr{F}_0$ of *special formulas* to be the least set $\mathscr{F}'$ with the following two properties:

(i) T, F, and every *special atomic formula* $t = s$ $(t, s \in \mathscr{T}_0)$ are in $\mathscr{F}'$.
(ii) $\neg\Phi$, *Surely* Φ, $(\Phi \vee \Psi)$, $(\Phi \wedge \Psi)$, and $(\Phi \Rightarrow \Psi)$ are in $\mathscr{F}'$ whenever $\Phi, \Psi \in \mathscr{F}'$.

As before, Theorem 3.2 implies

$$\|Surely\ \Phi\|_{\mathscr{S}} = \inf\{\|\Phi\|_{\mathscr{S}'} : \mathscr{S}' \text{ is a completion of } \mathscr{S}\}, \tag{17}$$

$$\|Possibly\ \Phi\|_{\mathscr{S}} = \sup\{\|\Phi\|_{\mathscr{S}'} : \mathscr{S}' \text{ is a completion of } \mathscr{S}\}. \tag{18}$$

The following lemma gives some useful properties of *surely*.

LEMMA 5.1

(*a*) *surely* $\langle i, A\rangle \approx_i \langle i, A\rangle$;
(*b*) *surely* $-\langle i, A\rangle \approx_i \langle i, D_i \backslash A\rangle$;
(*c*) *surely* $(t \cdot s) \approx_i$ *surely* $t \cdot$ *surely* s;
(*d*) *if* $t \approx_e s$, *then*
surely $t \approx_i$ *surely* s.

PROOF

(a) By Lemma 4.2(b),

$$\begin{aligned} surely\langle i, A\rangle &\approx_i \boxdot -\boxdot -\langle i, A\rangle \approx_i \boxdot -\langle i, D_i\backslash A\rangle \\ &\approx_i \langle i, D_i\backslash(D_i\backslash A)\rangle \approx_i \langle i, D_i\rangle. \end{aligned}$$

(b) We can prove (b) similarly, using Lemma 4.2(a) and (b).
(c) By (15), in any system $\mathscr{S}$ we have

$$\begin{aligned} \|surely\ (t\cdot s)\|_{\mathscr{S}} &= \bigcap\{\|t\cdot s\|_{\mathscr{S}'} : \mathscr{S}' \text{ is a completion of } \mathscr{S}\} \\ &= \bigcap\{\|t\|_{\mathscr{S}'} \cap \|s\|_{\mathscr{S}'} : \mathscr{S}' \text{ is a completion of } \mathscr{S}\} \\ &= \bigcap\{\|t\|_{\mathscr{S}'} : \mathscr{S}' \text{ is a completion of } \mathscr{S}\} \\ &\quad \cap \bigcap\{\|s\|_{\mathscr{S}'} : \mathscr{S}' \text{ is a completion of } \mathscr{S}\} \\ &= \|surely\ t\|_{\mathscr{S}} \cap \|surely\ s\|_{\mathscr{S}} \\ &= \|surely\ t\cdot surely\ s\|_{\mathscr{S}}. \end{aligned}$$

(d) $t \approx_e s$ means that $\|t\|_{\mathscr{S}'} = \|s\|_{\mathscr{S}'}$ for any complete system $\mathscr{S}'$. Hence, by (15),

$$\begin{aligned}\|\textit{surely } t\|_{\mathscr{S}} &= \bigcap\{\|t\|_{\mathscr{S}'}: \mathscr{S}' \text{ is a completion of } \mathscr{S}\}\\ &= \bigcap\{\|s\|_{\mathscr{S}'}: \mathscr{S}' \text{ is a completion of } \mathscr{S}\}\\ &= \|\textit{surely } s\|_{\mathscr{S}};\end{aligned}$$

that is, *surely* $t \approx_i$ *surely* s. □

By virtue of Lemma 5.1(d), if an occurrence of *surely* is within the scope of another occurrence, then the former can simply be deleted. Using this rule, we can easily transform any special term to a Boolean combination of descriptors and terms of the form *surely* s, s not containing *surely*. (Moreover, any reasonable term entering a database is, most probably, already of this form.)

In order to eliminate *surely* from the resulting term efficiently we need another lemma. First, however, we give some auxiliary definitions.

A term is said to be in *additive normal form* (ANF) if it is a sum of *primitive* terms of the form $\prod_{j\in J}\langle i_j, A_j\rangle$, where $i_p \neq i_q$ for $p \neq q$, and $\varnothing \neq A_j \neq D_{i_j}$ for all $j \in J$. Similarly, a term is in *multiplicative normal form* (MNF) if it is a product of *coprimitive* terms of the form $\sum_{j\in J}\langle i_j, A_j\rangle$, where, as before, all the attributes i_j are different and none of the descriptors reduces to 0 or 1. Of course, ANF and MNF are special cases of WANF and WMNF, respectively.

LEMMA 5.2. *If s is in MNF, then*

$$\textit{surely } s \approx_i s.$$

PROOF. Let s be of the form $\prod_i \sum_j \langle i_j, A_j\rangle$. Then

$$\begin{aligned}
\textit{surely } s &\approx_i \prod_i \textit{surely} \sum_j \langle i_j, A_j\rangle && \text{(by Lemma 5.1(c))}\\
&\approx_i \prod_i \boxdot\!-\!\boxdot\!-\sum_j \langle i_j, A_j\rangle && \text{(by definition of } \textit{surely})\\
&\approx_i \prod_i \boxdot\!-\!\boxdot\prod_j -\langle i_j, A_j\rangle && \text{(by de Morgan's Law)}\\
&\approx_i \prod_i \boxdot\!-\prod_j \boxdot\!-\langle i_j, A_j\rangle && \text{(by Lemma 5.1(c))}\\
&\approx_i \prod_i \boxdot\!-\prod_j \langle i_j, D_{i_j}\backslash A_j\rangle && \text{(by Lemma 4.2(b))}\\
&\approx_i \prod_i \boxdot\sum_j -\langle i_j, D_{i_j}\backslash A_j\rangle && \text{(by de Morgan's Law)}\\
&\approx_i \prod_i \sum_j \langle i_j, A_j\rangle && \text{(by Lemma 4.2(e))}\\
&\approx_i s. && \square
\end{aligned}$$

Now we are ready to summarize the simplified method of computing the value of an arbitrary special term t.

Step 1. Suppress nested occurrences of *surely* in t.

Step 2. Replace every subterm *surely* s by a term in MNF externally equivalent to s. (The process of transforming any simple term into MNF by using axioms in **B** is quite standard and the details are left to the reader; see also [11].)

Step 3. Transform the resulting term, which does not contain *surely*, into WMNF (see the proof of Lemma 4.3).

Step 4. Determine the value of the resulting WMNF term by using Theorem 4.3.

It is not difficult to give a complete set of axioms (involving—unlike **TB**—only special terms) for the internal interpretation of special terms.

Axioms for Special Terms Under Internal Interpretation. The set **S** of axioms consists of

S1. Substitutions of special terms into the axioms of Boolean algebra and the axioms:

(i) $surely\ (t \cdot s) = surely\ t \cdot surely\ s$,
(ii) $surely\ surely\ t = surely\ t$,
(iii) $surely\ 0 = 0$,
(iv) $surely\ 1 = 1$.

S2. The following axioms concerning descriptors:

(v) $\langle i, \varnothing \rangle = 0$,
(vi) $\langle i, D_i \rangle = 1$,
(vii) $\langle i, A \rangle \cdot \langle i, B \rangle = \langle i, A \cap B \rangle$ for all $i \in I$, $A, B \in \mathscr{B}_i$.

S3. The axiom

(viii) $$surely \sum_{p=1}^{k} \left(-\langle i_p, A_p \rangle + \sum_{q=1}^{n_p} \langle i_p, B_p^q \rangle \right) = \sum_{p=1}^{k} \langle i_p, (D_{i_p} \backslash A_p) \cup \bigcup_{q=1}^{n_p} B_p^q \rangle$$

for every positive integer k, every sequence of positive integers $n_1, \ldots, n_k$, every sequence of *distinct* attributes $i_1, \ldots, i_k$, and all $A_p, B_p^1, \ldots, B_p^{n_p} \in \mathscr{B}_{i_p}$, $1 \leq p \leq k$.

The completeness of **S** can be proved in the same way as the completeness of **TB**; what is essential is that **S** enables us to transform any special term into WMNF. We leave the proof to the reader.

As in the case of general terms, it may sometimes be convenient to use a dual axiom system, which is based on *possibly* as a primitive operation. The reader may easily verify that such an axiom system **S*** can be obtained from **S** by replacing axioms (i)–(iv) and (viii) by

(i)* $possibly\ (t + s) = possibly\ t + possibly\ s$,
(ii)* $possibly\ possibly\ t = possibly\ t$,
(iii)* $possibly\ 0 = 0$,
(iv)* $possibly\ 1 = 1$,
(viii)* $$possibly \prod_{p=1}^{k} \left(\langle i_p, A_p \rangle \cdot \prod_{q=1}^{n_p} -\langle i_p, B_p^q \rangle \right) = \prod_{p=1}^{k} -\langle i_p, (D_{i_p} \backslash A_p) \cup \bigcup_{q=1}^{n_p} B_p^q \rangle.$$

The above axiom system is especially useful when we transform special terms into weak additive normal form.

Example 5.1. Let us consider the following special term t:

$$\begin{aligned} &possibly\ (\langle \text{DEPT\#} = 4 \rangle \cdot \langle \text{NAME} = \text{BROWN} \rangle \\ &\qquad + surely\ (\langle \text{NAME} \neq \text{LIPSKI} \rangle \cdot \langle \text{DEPT\#} = 1 \rangle)) \\ &\cdot surely\ (\langle \text{SAL} < 15000 \rangle + \langle \#\text{CHILDREN} > 3 \rangle \\ &\qquad \cdot \langle \text{SAL IN } (10000, 20000) \rangle \cdot \langle \text{STATUS} = \text{MARRIED} \rangle). \end{aligned}$$

We show below the process of transforming t into WMNF. (We shall strictly follow the general method of transforming into WMNF described in this section, though in

our particular example there are places where the transformation can be done more efficiently.)

$$
\begin{aligned}
t \approx_i & -\textit{surely}- (\langle \text{DEPT}\# = 4\rangle \cdot \langle \text{NAME} = \text{BROWN}\rangle \\
& \quad + \langle \text{NAME} \neq \text{LIPSKI}\rangle \cdot \langle \text{DEPT}\# = 1\rangle) \\
& \cdot \textit{surely}\ ((\langle \text{SAL} < 15000\rangle + \langle \#\text{CHILDREN} > 3\rangle) \\
& \quad \cdot (\langle \text{SAL} < 15000\rangle + \langle \text{SAL IN } (10000, 20000)\rangle) \\
& \quad \cdot (\langle \text{SAL} < 15000\rangle + \langle \text{STATUS} = \text{MARRIED}\rangle)) \\
\approx_i & -(((\langle \text{DEPT}\# \neq 4\rangle + \langle \text{NAME} \neq \text{BROWN}\rangle) \\
& \quad \cdot (\langle \text{NAME} = \text{LIPSKI}\rangle + \langle \text{DEPT}\# \neq 1\rangle)) \\
& \cdot (\langle \text{SAL} < 15000\rangle + \langle \#\text{CHILDREN} > 3\rangle) \cdot \langle \text{SAL} < 20000\rangle \\
& \cdot (\langle \text{SAL} < 15000\rangle + \langle \text{STATUS} = \text{MARRIED}\rangle) \\
\approx_i & (-\langle \text{DEPT}\# \neq 4\rangle \cdot -\langle \text{NAME} \neq \text{BROWN}\rangle \\
& + -\langle \text{NAME} = \text{LIPSKI}\rangle \cdot -\langle \text{DEPT}\# \neq 1\rangle) \\
& \cdot (\langle \text{SAL} < 15000\rangle + \langle \#\text{CHILDREN} > 3\rangle) \cdot \langle \text{SAL} < 20000\rangle \\
& \cdot (\langle \text{SAL} < 15000\rangle + \langle \text{STATUS} = \text{MARRIED}\rangle) \\
\approx_i & (-\langle \text{DEPT}\# \neq 4\rangle + -\langle \text{NAME} = \text{LIPSKI}\rangle) \cdot -\langle \text{DEPT}\# \text{ IN } (2, 3, 5)\rangle \\
& \cdot -\langle \text{NAME} = \text{LIPSKI}\rangle \cdot (-\langle \text{NAME} \neq \text{BROWN}\rangle + -\langle \text{DEPT}\# \neq 1\rangle) \\
& \cdot (\langle \text{SAL} < 15000\rangle + \langle \#\text{CHILDREN} > 3\rangle) \cdot \langle \text{SAL} < 20000\rangle \\
& \cdot (\langle \text{SAL} < 15000\rangle + \langle \text{STATUS} = \text{MARRIED}\rangle).
\end{aligned}
$$

(In the transformation process we assume that $D_{\text{DEPT}\#} = \{1, 2, 3, 4, 5\}$.)

6. *Computing the Value of Special Formulas*

In this section we develop a method for determining $\|\Phi\|$ for any special formula Φ. It is interesting that this method has a combinatorial flavor and is quite different from that of computing the value of a special term. The main idea of our method can be explained using the following simple example.

Example 6.1. Suppose that three objects x, y, z are classified with respect to color. Assume that the color of no object is known, and consider the following two situations:

	I	II
possibly green objects	$A_1 = \{x, y\}$	$B_1 = \{x, y\}$
possibly red objects	$A_2 = \{x\}$	$B_2 = \{x\}$
possibly blue objects	$A_3 = \{x, y\}$	$B_3 = \{x, z\}$

(We do not exclude the possibility that an object is of another color than those listed above.) We may ask the following question: "Is it possible that all colors, that is, green, red and blue, are represented in our collection?" More formally, we ask for the value of the formula

$$\textit{Possibly}\ ((\textit{green} \neq 0) \wedge (\textit{red} \neq 0) \wedge (\textit{blue} \neq 0)). \tag{19}$$

It is easy to see that the answer for our question is "no" in case I and "yes" in case II. Indeed, in case I we have only the two objects x, y to represent three colors, while in case II it may be that x is red, y is green, and z is blue. In order to put this observation into more general terms, we need the following definition. A sequence $r_1, \ldots, r_n$ is said to be a *system of distinct representatives* (SDR) of a sequence of sets $S_1, \ldots, S_n$ if $r_i \in S_i$ for $1 \leq i \leq n$ and $r_i \neq r_j$ for $1 \leq i < j \leq n$. Coming back to our example, we see that the relevant difference between case I and case II is that there is no SDR for A_1, A_2, A_3, while there is one for B_1, B_2, B_3, namely, x, y, z. This SDR

provides an example of a possible completion of our information concerning the objects which makes formula (19) true. We may say that in a general situation of this type, involving any number of sets $S_1, \ldots, S_n$ corresponding to some mutually exclusive properties—call them "colors," the element r_i of an SDR of $S_1, \ldots, S_n$ plays the role of an object which "turns out to be of ith color."

The classical combinatorial theorem of Hall [4] asserts that an SDR for $S_1, \ldots, S_n$ exists if and only if

$$\left| \bigcup_{j \in J} S_j \right| \geq |J| \qquad \text{for every} \quad J \subseteq \{1, \ldots, n\}.$$

This condition is not very interesting from the algorithmic point of view, but efficient methods of testing for the existence of an SDR do exist. The best known algorithm is given by Hopcroft and Karp [5] in an equivalent formulation in terms of matchings in bipartite graphs. It may be useful to describe briefly this alternative formulation of the problem. Let $S_1, \ldots, S_n$ be subsets of $X = \{x_1, \ldots, x_m\}$. We construct a graph G with vertices corresponding to $S_1, \ldots, S_n, x_1, \ldots, x_m$, with an edge $\{S_i, x_j\}$ joining S_i and x_j, for all S_i, x_j such that $x_j \in S_i$. By a *matching* we mean any set of edges with the property that no two edges in this set are incident to a common vertex. It is clear that $r_1, \ldots, r_n$ is an SDR of $S_1, \ldots, S_n$ if and only if $\{S_1, r_1\}, \ldots, \{S_n, r_n\}$ is a matching in G.

Remark. The algorithm of Hopcroft and Karp *constructs* an SDR, while we are merely interested in its *existence*; it would be interesting to know whether testing for the existence of an SDR is strictly easier than constructing one.

Now we show how to decompose the problem of computing the value of an arbitrary special formula into some number of problems of the type described in Example 6.1. To this end we need some definitions and lemmas.

LEMMA 6.1. *For any formulas* Φ, Ψ,

(*a*) *Possibly* $(\Phi \vee \Psi) \approx_i$ *Possibly* $\Phi \vee$ *Possibly* Ψ;

(*b*) *if* $\Phi \approx_e \Psi$, *then Possibly* $\Phi \approx_i$ *Possibly* Ψ.

PROOF. The proof is analogous to the proof of Lemma 5.1(c) and (d). □

We may replace every occurrence of *Surely* by $\neg$*Possibly*$\neg$, and then, by Lemma 6.1(b), suppress all those occurrences of *Possibly*, *surely*, and *possibly* which are within the scope of *Possibly*. In this way we obtain a logical combination of special atomic formulas $t = s$ $(t, s \in \mathscr{T}_0)$ and formulas *Possibly* Φ, where Φ is simple (i.e., it does not contain *Surely*, *Possibly*, *surely*, or *possibly*). The value of any special atomic formula can easily be computed by the methods developed in the previous section; it is convenient to make use of the fact that

$$\|t = s\| = \begin{cases} T & \text{if} \quad \|t \cdot -s + s \cdot -t\| = \varnothing, \\ F & \text{otherwise.} \end{cases}$$

Determining $\|$*Possibly* $\Psi\|$, Ψ simple, is related to Example 6.1 and is much more difficult. We devote the rest of this section to showing how this can be done.

Definition 6.1

(i) A formula is *elementary* if it is of the form

$$(t_0 = 0) \wedge \bigvee_{k=1}^{n} (t_k \neq 0), \tag{20}$$

where $n \geq 0$, t_0 is in MNF, $t_1, \ldots, t_n$ are in ANF, and $t_p \cdot t_q \approx_e 0$ for $0 \leq p < q \leq n$.

(ii) A formula is in *special disjunctive normal form* (SDNF) if it is of the form

$$\bigwedge_{l \in L} \Psi,$$

where all Ψ_l's are elementary.

We treat formulas (20) lacking the atomic formula $t_0 = 0$ also as elementary; to be more formal, we could add in such a case a dummy atomic formula $0 = 0$.

We prove that every simple formula can be transformed into an externally equivalent formula in SDNF. Let us first notice that for any simple formula Φ there exists a finite collection of terms $s_1, \ldots, s_p$ (p depends on Φ) such that $s_i \cdot s_j \approx_e 0$ for $1 \leq i < j \leq p$, and every term occurring in Φ is externally equivalent to a sum of some number of terms s_i. We shall call such a collection a *set of atoms with respect to* Φ. One way to obtain $s_1, \ldots, s_p$ is the following. Let J be the set of attributes represented in Φ. For any $j \in J$, let

$$\mathscr{A}_j(\Phi) = \{A : \langle j, A \rangle \text{ occurs in } \Phi\}, \tag{21}$$

$$\mathscr{C}_j(\Phi) = \text{the set of all nonempty constituents of } \mathscr{A}_j(\Phi). \tag{22}$$

(Recall that a *constituent* of a family $\{A_1, \ldots, A_n\}$ of subsets of a set X is any set of the form $A_1^{\epsilon_1} \cap \cdots \cap A_n^{\epsilon_n}$ where $\epsilon_1, \ldots, \epsilon_n \in \{0, 1\}$, and $A_i^{\epsilon_i}$ denotes A_i if $\epsilon_i = 1$ and $X \setminus A_i$ if $\epsilon_i = 0$ (see Kuratowski and Mostowski [10, p. 21]).) In other words, $\mathscr{C}_j(\Phi)$ is the set of atoms of the Boolean algebra of subsets of D_j generated by $\mathscr{A}_j(\Phi)$. Now let us define

$$\mathscr{A}(\Phi) = \left\{ \prod_{j \in J} \langle j, A_j \rangle : \text{for every } j \in J,\ A_j \in \mathscr{C}_j(\Phi) \right\}.$$

It is easy to see that $\mathscr{A}(\Phi)$ is a set of atoms with respect to Φ. Sometimes it is convenient to consider instead of $\mathscr{C}_j(\Phi)$ a partition of D_j which is finer than the partition into nonempty constituents of $\mathscr{A}_j(\Phi)$. For instance, for a real-valued attribute it may be useful to consider a partition into disjoint intervals, even if not all constituents are intervals. For simplicity, we denote such a finer partition by $\mathscr{C}_j(\Phi)$, too.

THEOREM 6.1. *For every simple formula Φ there is a formula Ψ in SDNF such that $\Phi \approx_e \Psi$.*

Proof. We describe an effective procedure for transforming an arbitrary simple formula Φ into SDNF.

Step 1. Replace every atomic formula $t = s$ occurring in Φ by the externally equivalent formula $t \cdot -s + s \cdot -t = 0$.

Step 2. Using the propositional calculus axioms, transform the resulting formula into a disjunctive normal form $\bigvee_i \Phi_i$, where every Φ_i is of the form

$$\bigwedge_{k \in K^-} (u_k = 0) \wedge \bigwedge_{k \in K^+} (u_k \neq 0). \tag{23}$$

Now it is sufficient to show how to transform each such conjunction into SDNF.

Step 3. Using the equivalence

$$\bigwedge_{k \in K^-} (u_k = 0) \approx_e \left(\sum_{k \in K^-} u_k \right) = 0,$$

replace the first part of (23) by the atomic formula $t_0 = 0$, where t_0 is the result of transforming $\sum_{k \in K^-} u_k$ into MNF (for transforming into MNF, see [13]).

Step 4. Replace the resulting formula, that is,

$$(t_0 = 0) \wedge \mathop{\mathfrak{M}}_{k \in K^+} (u_k \neq 0), \tag{24}$$

by the (externally equivalent!) formula

$$(t_0 = 0) \wedge \mathop{\mathfrak{M}}_{k \in K^+} (u_k \cdot -t_0 \neq 0),$$

and then by

$$(t_0 = 0) \wedge \mathop{\mathfrak{M}}_{k \in K^+} (v_k \neq 0), \tag{25}$$

where v_k is the result of transforming $u_k \cdot -t_0$ into ANF.

Step 5. Let Θ be the second part of (25), that is,

$$\mathop{\mathfrak{M}}_{k \in K^-} (v_k \neq 0),$$

and let J be the set of attributes represented in Θ. For every $j \in J$, determine $\mathscr{C}_j(\Theta)$ (see (21), (22)). Since every v_k is in ANF, it is a sum of primitive terms of the form $\prod_{j \in P} \langle j, A_j \rangle$, $P \subseteq J$. Transform each such primitive term into $\prod_{j \in J} \langle j, A_j \rangle$ by adding factors $\langle j, D_j \rangle$ $(\approx_e 1)$ for every $j \in J \backslash P$. Now replace every factor $\langle j, A_j \rangle$ by the externally equivalent sum

$$\sum_{(C \subseteq A_j) \wedge (C \in \mathscr{C}_j(\Theta))} \langle i, C \rangle$$

(notice that $A_j = \cup\{C \in \mathscr{C}_j(\Theta) : C \subseteq A_j\}$ and that we have axiom (ii) of group B2), and then, using the distributive law, transform the resulting product of sums into a sum of atoms with respect to Θ. By applying the above transformation to every primitive term in v_j and then suppressing repeated summands, we transform v_j into a sum of distinct atoms with respect to Θ, say $\sum_{k \in M_j} s_k$.

Step 6. Using the equivalence

$$\left(\sum_{k \in M_j} s_k \right) \neq 0 \approx_e \mathop{\mathfrak{W}}_{k \in M_j} (s_k \neq 0),$$

transform (25) into

$$(t_0 = 0) \wedge \mathop{\mathfrak{M}}_{j \in K^+} \mathop{\mathfrak{W}}_{k \in M_j} (s_k \neq 0).$$

Applying the (logical) distributive law and then suppressing repeated formulas $s_k \neq 0$ within every conjunction, we ultimately arrive at a disjunction of elementary formulas, that is, SDNF. (It may then be useful to suppress repeated elementary formulas.) □

Example 6.2. Let Φ be the formula

$$\begin{aligned}
&(\langle \text{SEX} = \text{F} \rangle \neq \langle \text{AGE} \geq 25 \rangle) \\
&\quad \wedge [(\langle \text{AGE} < 30 \rangle \cdot \langle \text{SAL} > 30000 \rangle = 0) \wedge (\langle \text{SEX} = \text{M} \rangle = 0) \\
&\qquad \vee (\langle \text{AGE} \geq 25 \rangle \cdot \langle \text{SAL} < 15000 \rangle \cdot \langle \text{SEX} = \text{F} \rangle \neq 0) \\
&\qquad \wedge (\langle \text{SEX} = \text{M} \rangle \cdot \langle \text{SAL} \geq 15000 \rangle \neq 0)].
\end{aligned} \tag{26}$$

Below we show some of the stages of transforming Φ into SDNF:

$$\begin{aligned}
\Phi \approx_e\ & (\langle \mathrm{SEX} = \mathrm{F}\rangle \cdot \langle \mathrm{AGE} < 25\rangle + \langle \mathrm{SEX} = \mathrm{M}\rangle \cdot \langle \mathrm{AGE} \geq 25\rangle \neq 0) \\
& \quad \wedge (\langle \mathrm{AGE} < 30\rangle \cdot \langle \mathrm{SAL} > 30000\rangle = 0) \wedge (\langle \mathrm{SEX} = \mathrm{M}\rangle = 0) \\
& \vee (\langle \mathrm{SEX} = \mathrm{F}\rangle \cdot \langle \mathrm{AGE} < 25\rangle + \langle \mathrm{SEX} = \mathrm{M}\rangle \cdot \langle \mathrm{AGE} \geq 25\rangle \neq 0) \\
& \quad \wedge (\langle \mathrm{AGE} \geq 25\rangle \cdot \langle \mathrm{SAL} < 15000\rangle \cdot \langle \mathrm{SEX} = \mathrm{F}\rangle \neq 0) \\
& \quad \wedge (\langle \mathrm{SEX} = \mathrm{M}\rangle \cdot \langle \mathrm{SAL} \geq 15000\rangle \neq 0) \\
\approx_e\ & ((\langle \mathrm{AGE} < 30\rangle + \langle \mathrm{SEX} = \mathrm{M}\rangle) \cdot (\langle \mathrm{SAL} > 30000\rangle + \langle \mathrm{SEX} = \mathrm{M}\rangle) = 0) \\
& \quad \wedge ((\langle \mathrm{SEX} = \mathrm{F}\rangle \cdot \langle \mathrm{AGE} < 25\rangle \cdot \langle \mathrm{AGE} \geq 30\rangle \cdot \langle \mathrm{SEX} = \mathrm{F}\rangle \\
& \qquad + \langle \mathrm{SEX} = \mathrm{F}\rangle \cdot \langle \mathrm{AGE} < 25\rangle \cdot \langle \mathrm{SAL} \leq 30000\rangle \cdot \langle \mathrm{SEX} = \mathrm{F}\rangle \\
& \qquad + \langle \mathrm{SEX} = \mathrm{M}\rangle \cdot \langle \mathrm{AGE} \geq 25\rangle \cdot \langle \mathrm{AGE} \geq 30\rangle \cdot \langle \mathrm{SEX} = \mathrm{F}\rangle \\
& \qquad + \langle \mathrm{SEX} = \mathrm{M}\rangle \cdot \langle \mathrm{AGE} \geq 25\rangle \cdot \langle \mathrm{SAL} \leq 30000\rangle \cdot \langle \mathrm{SEX} = \mathrm{F}\rangle) \neq 0) \\
& \vee (\langle \mathrm{SEX} = \mathrm{F}\rangle \cdot \langle \mathrm{AGE} < 25\rangle \neq 0) \\
& \quad \wedge (\langle \mathrm{AGE} \geq 25\rangle \cdot \langle \mathrm{SAL} < 15000\rangle \cdot \langle \mathrm{SEX} = \mathrm{F}\rangle \neq 0) \\
& \quad \wedge (\langle \mathrm{SEX} = \mathrm{M}\rangle \cdot \langle \mathrm{SAL} \geq 15000\rangle \neq 0) \\
& \vee (\langle \mathrm{SEX} = \mathrm{M}\rangle \cdot \langle \mathrm{AGE} \geq 25\rangle \neq 0) \\
& \quad \wedge (\langle \mathrm{AGE} \geq 25\rangle \cdot \langle \mathrm{SAL} < 15000\rangle \cdot \langle \mathrm{SEX} = \mathrm{F}\rangle \neq 0) \\
& \quad \wedge (\langle \mathrm{SEX} = \mathrm{M}\rangle \cdot \langle \mathrm{SAL} \geq 15000\rangle \neq 0).
\end{aligned}$$

The first two parts of the above disjunction are already elementary formulas (strictly speaking, the first one becomes an elementary formula after some simple reductions). Let Θ be the third part. We have

$$\begin{aligned}
& \mathscr{C}_{\mathrm{AGE}}(\Theta) = \{(0, 25), [25, \infty)\}, \\
& \mathscr{C}_{\mathrm{SAL}}(\Theta) = \{[0, 15000), [15000, \infty)\}, \\
& \mathscr{C}_{\mathrm{SEX}}(\Theta) = \{\{F\}, \{M\}\}, \\
& \quad \Theta \approx_e (\langle \mathrm{SEX} = \mathrm{M}\rangle \cdot \langle \mathrm{AGE} \geq 25\rangle \cdot (\langle \mathrm{SAL} < 15000\rangle + \langle \mathrm{SAL} \geq 15000\rangle) \neq 0) \\
& \qquad \wedge (\langle \mathrm{AGE} \geq 25\rangle \cdot \langle \mathrm{SAL} < 15000\rangle \cdot \langle \mathrm{SEX} = \mathrm{F}\rangle \neq 0) \\
& \qquad \wedge (\langle \mathrm{SEX} = \mathrm{M}\rangle \cdot \langle \mathrm{SAL} \geq 15000\rangle \\
& \qquad\quad \cdot (\langle \mathrm{AGE} < 25\rangle + \langle \mathrm{AGE} \geq 25\rangle) \neq 0).
\end{aligned}$$

After applying the distributive law and performing some simple reductions, we finally obtain the SDNF of Φ:

$$\begin{aligned}
& ((\langle \mathrm{AGE} < 30\rangle + \langle \mathrm{SEX} = \mathrm{M}\rangle) \cdot (\langle \mathrm{SAL} > 30000\rangle + \langle \mathrm{SEX} = \mathrm{M}\rangle) = 0) \\
& \quad \wedge (\langle \mathrm{SEX} = \mathrm{F}\rangle \cdot \langle \mathrm{AGE} < 25\rangle \cdot \langle \mathrm{SAL} \leq 30000\rangle \neq 0) \\
& \vee (\langle \mathrm{SEX} = \mathrm{F}\rangle \cdot \langle \mathrm{AGE} < 25\rangle \neq 0) \\
& \quad \wedge (\langle \mathrm{AGE} \geq 25\rangle \cdot \langle \mathrm{SAL} < 15000\rangle \cdot \langle \mathrm{SEX} = \mathrm{F}\rangle \neq 0) \\
& \quad \wedge (\langle \mathrm{SEX} = \mathrm{M}\rangle \cdot \langle \mathrm{SAL} \geq 15000\rangle \neq 0) \\
& \vee (\langle \mathrm{SEX} = \mathrm{M}\rangle \cdot \langle \mathrm{AGE} \geq 25\rangle \cdot \langle \mathrm{SAL} < 15000\rangle \neq 0) \\
& \quad \wedge (\langle \mathrm{AGE} \geq 25\rangle \cdot \langle \mathrm{SAL} < 15000\rangle \cdot \langle \mathrm{SEX} = \mathrm{F}\rangle \neq 0) \\
& \quad \wedge (\langle \mathrm{SEX} = \mathrm{M}\rangle \cdot \langle \mathrm{SAL} \geq 15000\rangle \cdot \langle \mathrm{AGE} < 25\rangle \neq 0) \\
& \vee (\langle \mathrm{SEX} = \mathrm{M}\rangle \cdot \langle \mathrm{AGE} \geq 25\rangle \cdot \langle \mathrm{SAL} < 15000\rangle \neq 0) \\
& \quad \wedge (\langle \mathrm{AGE} \geq 25\rangle \cdot \langle \mathrm{SAL} < 15000\rangle \cdot \langle \mathrm{SEX} = \mathrm{F}\rangle \neq 0) \\
& \quad \wedge (\langle \mathrm{SEX} = \mathrm{M}\rangle \cdot \langle \mathrm{SAL} \geq 15000\rangle \cdot \langle \mathrm{AGE} \geq 25\rangle \neq 0) \\
& \vee (\langle \mathrm{SEX} = \mathrm{M}\rangle \cdot \langle \mathrm{AGE} \geq 25\rangle \cdot \langle \mathrm{SAL} \geq 15000\rangle \neq 0) \\
& \quad \wedge (\langle \mathrm{AGE} \geq 25\rangle \cdot \langle \mathrm{SAL} < 15000\rangle \cdot \langle \mathrm{SEX} = \mathrm{F}\rangle \neq 0) \\
& \quad \wedge (\langle \mathrm{SEX} = \mathrm{M}\rangle \cdot \langle \mathrm{SAL} \geq 15000\rangle \cdot \langle \mathrm{AGE} < 25\rangle \neq 0) \\
& \vee (\langle \mathrm{SEX} = \mathrm{M}\rangle \cdot \langle \mathrm{AGE} \geq 25\rangle \cdot \langle \mathrm{SAL} \geq 15000\rangle \neq 0) \\
& \quad \wedge (\langle \mathrm{AGE} \geq 25\rangle \cdot \langle \mathrm{SAL} < 15000\rangle \cdot \langle \mathrm{SEX} = \mathrm{F}\rangle \neq 0). \qquad (27)
\end{aligned}$$

Notice that in order to be more efficient, we did not exactly follow the general pattern of transformation described in the proof of Theorem 6.1.

All we need now is a method for computing the value of *Possibly* Φ, Φ elementary.

Indeed, let us transform a simple formula Ψ into an externally equivalent formula in SDNF, say $\bigvee_{l\in L}\Psi_l$, where the Ψ_l are elementary. By Lemma 6.1,

$$\textit{Possibly } \Psi \approx_i \textit{Possibly } \bigvee_{l\in L}\Psi_l \approx_i \bigvee_{l\in L} \textit{Possibly } \Psi_l.$$

The last step is provided by the next theorem.

THEOREM 6.2. *Let Φ be an elementary formula of the form*

$$(t_0 = 0) \wedge \bigwedge_{k=1}^{n} (t_k \neq 0). \tag{28}$$

Then $\| \textit{Possibly } \Phi \|_{\mathscr{S}} = T$ if and only if the following two conditions are satisfied:

(i) $\| \textit{surely } t_0 \|_{\mathscr{S}} = \varnothing$;

(ii) the sequence $\| \textit{possibly } t_1 \|_{\mathscr{S}}, \ldots, \| \textit{possibly } t_n \|_{\mathscr{S}}$ has an SDR.

PROOF. If $\| \textit{Possibly } \Phi \|_{\mathscr{S}} = T$, then by (18) there is a completion $\mathscr{S}' \geq \mathscr{S}$ such that $\| \Phi \|_{\mathscr{S}'} = T$; that is, $\| t_0 \|_{\mathscr{S}'} = \varnothing$ and $\| t_k \|_{\mathscr{S}'} \neq \varnothing$, $1 \leq k \leq n$. By (15) it follows that $\| \textit{surely } t_0 \|_{\mathscr{S}} = \varnothing$, and since $t_p \cdot t_q \approx_e 0$ for $1 \leq p < q \leq n$, the sets $\| t_1 \|_{\mathscr{S}'}, \ldots, \| t_n \|_{\mathscr{S}'}$ are mutually disjoint, so that we may choose distinct elements $x_1 \in \| t_1 \|_{\mathscr{S}'}, \ldots, x_n \in \| t_n \|_{\mathscr{S}'}$. But (16) implies $\| t_k \|_{\mathscr{S}'} \subseteq \| \textit{possibly } t_k \|_{\mathscr{S}}$; hence $x_1, \ldots, x_n$ is an SDR of the sequence $\| \textit{possibly } t_1 \|_{\mathscr{S}}, \ldots, \| \textit{possibly } t_n \|_{\mathscr{S}}$.

Conversely, suppose that $\| \textit{surely } t_0 \|_{\mathscr{S}} = \varnothing$ and that $x_1, \ldots, x_n$ is an SDR of the sequence $\| \textit{possibly } t_1 \|_{\mathscr{S}}, \ldots, \| \textit{possibly } t_n \|_{\mathscr{S}}$. Then by (15) and (16) there exist completions $\mathscr{S}_0, \mathscr{S}_1, \ldots, \mathscr{S}_n$ such that $\| t_0 \|_{\mathscr{S}_0} = 0$, $x_k \in \| t_k \|_{\mathscr{S}_k}$, $1 \leq k \leq n$. Let $(\beta_i^k)_{i\in I}$, $0 \leq k \leq n$, correspond to $\mathscr{S}_k$, $0 \leq k \leq n$, respectively. We define a completion $\mathscr{S}'$ by

$$\beta_i(x) = \begin{cases} \beta_i^k(x) & \text{if } x = x_k \text{ for some } k, \quad 1 \leq k \leq n, \\ \beta_i^0(x) & \text{otherwise.} \end{cases}$$

It is easy to see that $x_k \in \| t_k \|_{\mathscr{S}'}$, $1 \leq k \leq n$, and $\| t_0 \|_{\mathscr{S}'} = \varnothing$ ($x_k \notin \| t_0 \|_{\mathscr{S}'}$ since $x_k \in \| t_k \|_{\mathscr{S}'}$ and $t_0 \cdot t_k \approx_e 0$). Consequently, $\| \Phi \|_{\mathscr{S}'} = T$, that is, $\| \textit{Possibly } \Phi \|_{\mathscr{S}} = T$. □

Notice that in the proof we did not make use of the fact that t_0 was in MNF and that the t_k, $1 \leq k \leq n$, were in ANF. In fact, these assumptions in the definition of an elementary formula (see Definition 6.1(i)) were made only to make computing the value easier: if t_0 is in MNF, then by Lemma 5.2,

$$\| \textit{surely } t_0 \| = \| t_0 \|; \tag{29}$$

similarly, if t_k is in ANF, say $\sum_p \prod_q \langle i_{pq}, A_{pq} \rangle$, then

$$\begin{aligned} \| \textit{possibly } t_k \| &= \left\| -\textit{surely} \prod_p \sum_q \langle i_{pq}, D_{i_{pq}} \backslash A_{pq} \rangle \right\| \\ &= \left\| -\prod_p \sum_q \langle i_{pq}, D_{i_{pq}} \backslash A_{pq} \rangle \right\| \\ &= \left\| \sum_p \prod_q -\langle i_{pq}, D_{i_{pq}} \backslash A_{pq} \rangle \right\|. \end{aligned} \tag{30}$$

Example 6.3. Consider a very simple system represented by the following table:

object	AGE	SAL	SEX
x_1	(20, 40)	[0, ∞)	{M}
x_2	{35}	[30000, 40000)	{F, M}
x_3	(30, 40)	[20000, 30000]	{F}
x_4	[20, 30)	(10000, 20000)	{F}

We shall compute $\| \textit{Possibly } \Phi \|$, where Φ is given by (26). First we transform Φ into SDNF, which yields formula (27). This formula is a disjunction of six elementary formulas, $\Phi_1 \vee \cdots \vee \Phi_6$, where Φ_i is of the form $(t_{i0} = 0) \wedge \bigwedge_{k=1}^{n_i}(t_{ik} \neq 0)$ (t_{i0} may be 0). Using (29), (30), and Theorem 4.3, we obtain

$\| \textit{Possibly } \Phi_1 \| = F$, since $\| \textit{surely } t_{10} \| = \{x_1\} \neq \varnothing$,
$\| \textit{Possibly } \Phi_2 \| = F$, since no SDR exists for
$\| \textit{possibly } t_{21} \| = \{x_4\}$, $\| \textit{possibly } t_{22} \| = \{x_4\}$, $\| \textit{possibly } t_{23} \| = \{x_1, x_2\}$,
$\| \textit{Possibly } \Phi_3 \| = F$, since no SDR exists for
$\| \textit{possibly } t_{31} \| = \{x_1\}$, $\| \textit{possibly } t_{32} \| = \{x_4\}$, $\| \textit{possibly } t_{33} \| = \{x_1\}$,
$\| \textit{Possibly } \Phi_4 \| = T$, since x_1, x_4, x_3 is an SDR for
$\| \textit{possibly } t_{41} \| = \{x_1\}$, $\| \textit{possibly } t_{42} \| = \{x_4\}$, $\| \textit{possibly } t_{43} \| = \{x_1, x_2\}$.

Consequently, $\| \textit{Possibly } \Phi \| = T$ (we need not consider Φ_5 and Φ_6).

It is easy to see that the length of the SDNF may, in general, grow exponentially when the length of a formula gets large. Notice, however, that any straightforward method of evaluating $\| \textit{Possibly } \Phi \|_{\mathscr{S}}$ based on enumerating all completions of $\mathscr{S}$ would be incomparably worse, since the number of such completions is, in general, an exponential function of the number of objects. (Strictly speaking, the number of completions may be infinite when an attribute domain is infinite. However, we do not need the exact value of an attribute j in a completion; it is sufficient to know the subset $C \in \mathscr{C}_j(\Phi)$ to which it belongs.) For obvious reasons, our simple example could not illustrate the fact that in a real database the number of objects is usually several orders of magnitude bigger than the length of a query.

Some ways of improving the efficiency of our method of evaluating $\| \textit{Possibly } \Phi \|$ are listed below:

(1) In order to transform (25) into SDNF we may repeatedly apply the equivalence

$$(v_k \neq 0) \wedge (v_l \neq 0) \approx_e (v_k \cdot v_l \neq 0) \vee ((v_k \neq 0) \wedge (v_k \cdot -v_l \neq 0))$$

and the distributive law to formulas $(v_k \neq 0)$, $(v_l \neq 0)$ such that $v_k \cdot v_l \not\approx_e 0$. The resulting SDNF contains, in general, fewer elementary formulas. In particular, we skip Steps 5 and 6 whenever (25) is already an elementary formula.

(2) We need not transform (24) into SDNF if we find that $\| \textit{surely } t_0 \| \neq \varnothing$ (the value of $\| \textit{Possibly } \Psi \|$, where Ψ is given by (24), is then F). After having generated an elementary formula Φ_i we may evaluate $\| \textit{Possibly } \Phi_i \|$, and if it is T, we need not continue the transformation process (the value of the whole formula *Possibly* Φ is then T).

(3) Testing for the existence of an SDR can be simplified by using the following simple combinatorial facts:

(a) If $S_k = \varnothing$ for some k, then no SDR exists for $S_1, \ldots, S_n$.
(b) If $S_k \setminus (S_1 \cup \cdots \cup S_{k-1} \cup S_{k+1} \cup \cdots \cup S_n) \neq \varnothing$, then

$S_1, \ldots, S_n$ has an SDR $\Leftrightarrow S_1, \ldots, S_{k-1}, S_{k+1}, \ldots, S_n$ has an SDR.

(c) If $|S_k| \geq n$, then

$S_1, \ldots, S_n$ has an SDR $\Leftrightarrow S_1, \ldots, S_{k-1}, S_{k+1}, \ldots, S_n$ has an SDR.

(a) corresponds to $\| \textit{possibly } t_k \|_{\mathscr{S}} = \varnothing$, and (b) to $\| \textit{surely } t_k \|_{\mathscr{S}} \neq \varnothing$, since $t_k \cdot t_i \approx_e 0$ implies

$$\| \textit{possibly } t_k \| \setminus \| \textit{possibly } t_i \| \supseteq \| \textit{surely } t_k \| \setminus \| -\textit{surely} -t_i \|$$
$$= \| \textit{surely } t_k \| \cap \| \textit{surely} -t_i \| = \| \textit{surely } t_k \cdot -t_i \| = \| \textit{surely } t_k \| \neq \varnothing.$$

(4) If t_0 in 0 is an elementary formula Ψ and $\| \textit{Possibly } \Psi \| = T$ for a subset of our collection of objects, then $\| \textit{Possibly } \Psi \| = T$ for the whole collection. It is not necessarily so when t_0 is not 0; then the existence of an SDR for $\| \textit{possibly } t_1 \|, \ldots, \| \textit{possibly } t_n \|$ is preserved under adding new objects, but the condition $\| \textit{surely } t_0 \| = \emptyset$ may be violated.

By the last remark, we may infer that the value of *Possibly* Φ, where Φ given by (26), is T in any system which—when represented by a table—contains the four rows from Example 6.3.

Example 6.4. Let Φ be the formula

$$(((\langle \text{HAIR} = \text{FAIR} \rangle \cdot \textit{possibly } \langle \text{SEX} = \text{F} \rangle) = 0) \\ \wedge \textit{Surely } (\langle \text{SAL} > 50000 \rangle \cdot \langle \text{TAX} < 5000 \rangle = 0).$$

The first part of this formula can be transformed into

$$\langle \text{HAIR} = \text{FAIR} \rangle \cdot -\langle \text{SEX} = \text{M} \rangle = 0$$

(the left-hand side is in both WANF and WMNF, which enables us to easily compute its value in any system). In order to evaluate the second part of our formula, we might exploit the general method based on transforming into SDNF. It is, however, not necessary in our simple example (neither is it necessary in most of the queries likely to arise in practice). Instead, we may use the first of the following four equivalencies:

$$\begin{aligned} \textit{Surely } (t = 0) &\approx_i (\textit{possibly } t = 0), \\ \textit{Surely } (t \neq 0) &\approx_i (\textit{surely } t \neq 0), \\ \textit{Possibly } (t = 0) &\approx_i (\textit{surely } t = 0), \\ \textit{Possibly } (t \neq 0) &\approx_i (\textit{possibly } t \neq 0), \end{aligned}$$

(t is an arbitrary special term; the easy proof of these equivalencies is left to the reader). We have

$$\textit{possibly } (\langle \text{SAL} > 50000 \rangle \cdot \langle \text{TAX} < 5000 \rangle) \approx_i -\langle \text{SAL} \leq 50000 \rangle \cdot -\langle \text{TAX} \geq 5000 \rangle$$

(see axiom (viii)*), and consequently our formula is transformed into the form

$$(\langle \text{HAIR} = \text{FAIR} \rangle \cdot -\langle \text{SEX} = \text{M} \rangle = 0) \\ \wedge (-\langle \text{SAL} \leq 50000 \rangle \cdot -\langle \text{TAX} \geq 5000 \rangle = 0),$$

which easily lends itself to evaluation in any system.

In the above example we have eliminated *Surely* from our formula. This is, however, not always possible. The reason for this impossibility is, very roughly speaking, the fact that the existence of an SDR is not a "Boolean" property; there may be two sequences, $S_1, \ldots, S_n$ of subsets of X and $T_1, \ldots, T_n$ of subsets of Y, such that

$$S_1^{\epsilon_1} \cap \cdots \cap S_n^{\epsilon_n} = \emptyset \Leftrightarrow T_1^{\epsilon_1} \cap \cdots \cap T_n^{\epsilon_n} = \emptyset$$

for all $\epsilon_1, \ldots, \epsilon_n \in \{0, 1\}$, yet $S_1, \ldots, S_n$ has an SDR, while $T_1, \ldots, T_n$ does not have any SDR. This is so, for instance, in the case of the sequences $\{x, y\}$, $\{x, y\}$ and $\{x\}$, $\{x\}$. It is easy to see that formulas not containing *Surely* express "Boolean" properties of the values of special terms. On the other hand, we know (see Theorem 6.2) that there are special formulas of the form *Possibly* Ψ which express the existence of an SDR—a "non-Boolean" property—for the values of some special terms. Of course, no such formula can be replaced by any (internally) equivalent formula not

containing *Surely*. For example, let us consider a system with $\beta_{\mathrm{SEX}}(x) = \{\mathrm{F}, \mathrm{M}\} = D_{\mathrm{SEX}}$ for all $x \in X$, and let Φ be the formula *Possibly* $((\langle \mathrm{SEX} = \mathrm{F}\rangle \neq 0) \wedge (\langle \mathrm{SEX} = \mathrm{M}\rangle \neq 0))$. Of course, if $|X| \leq 1$, then $\|\Phi\| = F$, and if $|X| \geq 2$, then $\|\Phi\| = T$. However, the value of any formula not containing *Surely* is the same in both cases. Indeed, the value of any special term is either $\emptyset$ in both cases or X in both cases, and hence the value of any atomic formula is the same in both cases.

To conclude this section, let us mention that the values $\|$ *surely* $t\|$, $\|$ *possibly* $t\|$, $\|$ *Surely* $\Phi\|$, and $\|$ *Possibly* $\Phi\|$ (for simple t and Φ) were denoted in [13] by $\|t\|_*$, $\|t\|^*$, $\|\Phi\|_*$, and $\|\Phi\|^*$, respectively ($\|\cdot\|_*$ and $\|\cdot\|^*$ were called the *lower value* and the *upper value*, respectively). Our algorithm for computing $\|$ *Possibly* $\Phi\|$ gives a method for determining $\|\Phi\|^*$ (and $\|\Phi\|_*$, since $\|\Phi\|_* = \neg\|\neg\Phi\|^*$), a problem which was left open in [13].

7. *Other Interpretations of Queries*

There us another approach to semantics of queries in an incomplete information system, based on the theory of pseudo-Boolean algebras (PBAs) and intuitionistic logic. (The relation of PBAs to intuitionistic logic is exactly the same as the relation of TBAs to modal logic S4; see [3, 15].) We describe only the pseudo-Boolean approach to interpreting terms (intuitionistic interpretation of formulas is considered in [11]). In the pseudo-Boolean approach we consider only simple queries, and we treat "$-$" as a "strong" negation, that is, $-t$ is understood, roughly speaking, as the set of objects *known not to have* property t, instead of just *not known to have* property t. Also the interpretation of "$\rightarrow$" is "strong." The formal definition of the "pseudo-Boolean value" of a simple term t—denote it by $|t|_{\mathscr{S}}$—can be obtained by changing (iii) and (vi) in Definition 3.1 to

$$|-t|_{\mathscr{S}} = \{x \in X: \text{for every } \mathscr{S}' \geq \mathscr{S},\ x \notin |t|_{\mathscr{S}'}\},$$
$$|t \rightarrow s|_{\mathscr{S}} = \{x \in X: \text{for every } \mathscr{S}' \geq \mathscr{S},\ x \notin |t|_{\mathscr{S}'} \text{ or } x \in |s|_{\mathscr{S}'}\}$$

(and deleting (vii)). It can easily be shown that for any simple term t,

$$|t|_{\mathscr{S}} = \|\tau(t)\|_{\mathscr{S}},$$

where $\tau(t)$ is defined inductively by

(i) $\tau(0) = 0, \quad \tau(1) = 1,$
(ii) $\tau(t + s) = \tau(t) + \tau(s),$
(iii) $\tau(t \cdot s) = \tau(t) \cdot \tau(s),$
(iv) $\tau(-t) = \boxdot -\tau(t),$
(v) $\tau(t \rightarrow s) = \boxdot(\tau(t) \rightarrow \tau(s)).$

It is also not difficult to prove that

$$\|\textit{surely } t\|_{\mathscr{S}} = |--t|_{\mathscr{S}},$$
$$\|\textit{possibly } t\|_{\mathscr{S}} = X \setminus |-t|_{\mathscr{S}}.$$

Similar relations between $\|\cdot\|$ and $|\cdot|$ exist for formulas. These relations reflect the fact that the set of open elements of any TBA (a is *open* if $a = \mathbb{I}a$) forms a PBA, and they are connected with the well-known interpretation of intuitionistic logic within the modal logic S4 (see [3, 15]). It may be noted that there exists a similarity between a "pseudo-Boolean value" and the Kripke models for the intuitionistic logic (Kripke [9]; see also Fitting [3]). It can also be shown that $|\cdot|_{\mathscr{S}}$ coincides with the interpretation of terms defined—in a quite different way—by Jaegermann [7].

In the internal interpretation of queries presented in this paper we consider, for any incomplete system $\mathcal{S}$, the set of *all* extensions of $\mathcal{S}$. Since $\mathcal{S}$ represents an incomplete knowledge about a reality described by a completion $\mathcal{S}^*$ of $\mathcal{S}$, not all of these extensions are really possible. The only completions accessible in reality are those consistent with $\mathcal{S}^*$, that is, those of which $\mathcal{S}^*$ is a completion. Hence, we may consider a different approach, where the partially ordered set of all extensions of $\mathcal{S}$ is replaced by

$$[\mathcal{S}, \mathcal{S}^*] = \{\mathcal{S}' : \mathcal{S} \preceq \mathcal{S}' \preceq \mathcal{S}^*\}.$$

Of course, in general we do not know $\mathcal{S}^*$. However, there are formulas, such as

$$\boxdot\diamond t = \diamond\boxdot t, \qquad \Box\diamond\Phi \Leftrightarrow \diamond\Box\Phi,$$

which are true for any $\mathcal{S}$ and for any completion $\mathcal{S}^*$ of $\mathcal{S}$. (The interpretation of a term and of a formula is now the same as in Definition 3.1, but with "for every $\mathcal{S}' \succeq \mathcal{S}\ldots$" replaced by "for every $\mathcal{S}' \in [\mathcal{S}, \mathcal{S}^*]\ldots$.") These universally valid formulas define a logic, quite different from the one inherent in our internal interpretation described in Section 4. Notice that this logic describes the process of increasing the user's (system's) knowledge, as seen by an observer who has complete information about both the system ($\mathcal{S}$) and reality ($\mathcal{S}^*$).

8. *Conclusions*

Following a treatment of more elementary topics in [13], we have given a thorough treatment of the internal interpretation of queries. It has turned out that this interpretation leads in a natural way to the notion of a topological Boolean algebra and to a modal logic related to S4. These notions have been shown to play the same role for the internal interpretation as Boolean algebras and classical logic in the case of external interpretation. We have presented a complete axiom system for internally equivalent transformations of terms and a method for computing the internal interpretation for arbitrary terms and for a broad class of formulas including most formulas of practical interest.

The number of steps performed by our algorithms may, in the worst case, be an exponential function of the length of the query. However, they seem to be of practical interest for real world queries that are likely to be submitted to a database.

There are a number of interesting problems which remain open. One such problem is to investigate in more detail the logic involved in the internal interpretation of formulas. More specifically, it is not known to the author whether this logic is decidable (it may be noted that the results of Section 4 can easily be used to develop a decision procedure for *atomic* formulas). Another question is whether there is a simple axiom system for this logic. Both problems are open even for the sublanguage considered in Section 5.

Note added in proof. The logic involved in the internal equivalence of formulas has recently been shown to be decidable; see H. Ono and A. Nakamura, Decidability results on a query language for data bases with incomplete information, Proc. 9th Int. Symp. on Mathematical Foundations of Computer Science, P. Dembinski, Ed., Springer-Verlag, Berlin, 1980, pp. 452–459.

ACKNOWLEDGMENTS. I am indebted to many individuals for their assistance at various stages of the development of this paper. My special thanks are due to W. Marek and Z. Pawlak, who encouraged me to work on the problem of incomplete

information. I have benefited greatly from discussions with M. Jaegermann, J. Łoś, C. Rauszer, and K. Segerberg. The suggestions of Dave Harrel resulted in considerable improvements in the presentation of the results.

REFERENCES

1. CODD, E.F. A relational model of data for large shared data banks. *Commun. ACM 13*, 6 (June 1970), 377–387.
2. CODD, E.F. Understanding relations (Installment #7). *FDT Bulletin of ACM-SIGMOD 7*, 3-4 (1975), 23–28.
3. FITTING, M.C. *Intuitionistic Logic, Model Theory and Forcing.* North-Holland, Amsterdam, 1969.
4. HALL, P. On representatives of subsets. *J. London Math. Soc. 10* (1935), 26–30.
5. HOPCROFT, J.E., AND KARP, R.M. An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs. *SIAM J. Comput. 2* (1973), 225–231.
6. HUGHES, G.E., AND CRESSWELL, M.J. *An Introduction to Modal Logic.* Methuen and Company, London, 1972.
7. JAEGERMANN, M. Information storage and retrieval systems with incomplete information I, II. *Fundamenta Inform. 2* (1978), 17–41, *2* (1979), 141–166.
8. KRIPKE, S.A. Semantical analysis of modal logic I. *Z. Math. Logik Grundlagen Math. 9* (1963), 67–96.
9. KRIPKE, S.A. Semantical analysis of intuitionistic logic. In *Formal Systems and Recursive Functions*, J.N. Crossley and M.A.E. Dummett, Eds., North-Holland, Amsterdam, 1965, pp. 92–129.
10. KURATOWSKI, K., AND MOSTOWSKI, A. *Set Theory.* Polish Scientific Publishers, Warsaw 1976.
11. LIPSKI, W. Informational systems with incomplete information. Proc. 3rd Int. Symp. on Automata, Languages and Programming, Edinburgh, 1976; S. Michaelson and R. Milner, Eds., Edinburgh University Press, Edinburgh, 1976, pp. 120–130.
12. LIPSKI, W. On the logic of incomplete information. Proc. 6th Int. Symp. on Mathematical Foundations of Computer Science, Tatranska Lomnica, Czechoslovakia, Sept. 5–9, 1977; T. Gruska, Ed., Springer-Verlag, Berlin, 1977, pp. 374–381.
13. LIPSKI, W. On semantic issues connected with incomplete information databases. *ACM Trans. Database Syst. 4*, 3 (Sept. 1979), 262–296.
14. MAREK, W., AND PAWLAK, Z. Information storage and retrieval systems: Mathematical foundations. *Theoret. Comput. Sci. 1* (1976), 331–354.
15. RASIOWA, H., AND SIKORSKI, R. *The Mathematics of Metamathematics.* Polish Scientific Publishers, Warsaw, 1963.

RECEIVED FEBRUARY 1979; REVISED SEPTEMBER 1979; ACCEPTED JANUARY 1980